THE BELL SYSTEM
TECHNICAL JOURNAL
DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 52 September 1973 Number 7

A Coding Theorem for Multiple Access Channels With Correlated Sources

By DAVID SLEPIAN and JACK KEIL WOLF†

(Manuscript received February 22, 1973)

A communication system is studied in which two users communicate with one receiver over a common discrete memoryless channel. The information to be transmitted by the users may be correlated. Their information rates are described by a point in a suitably defined three-dimensional rate space.

A point in this rate space is called admissible *if there exist coders and decoders for the channel that permit the users to transmit information over it at the corresponding rates with arbitrarily small error probability. The closure of the set of all admissible rate points is called the* capacity region, $\mathcal{C}$, *and is the natural generalization of channel capacity to this situation.*

In this paper we show that $\mathcal{C}$, *which depends only on the channel, is convex and we give formulas to determine it exactly. Several simple channels are treated in detail and their capacity regions given explicitly.*

I. INTRODUCTION

The mathematical theory of communication has been concerned, for the most part, with the reliable transmission of information from a single information source to a single user. An extensive literature exists

† J. K. Wolf is Professor of Electrical Engineering at the University of Massachusetts, Amherst, Mass. Partial support for his research on this paper was furnished by the Air Force Office of Scientific Research under contract F-44620-72-C-0085.

on this problem: tbe basic concepts are contained in the classic papers of Shannon.[1]

In this work we consider tbe case in which messages from a *set* of information sources are communicated over a common channel to a single receiver. We impose constraints on tbe encoding techniques which can be employed.

A precise formulation of the problem is presented in a subsequent section. Here we describe in less mathematical terms the type of problem considered.

A particular multiple access communication channel with two inputs and one output is shown in Fig. 1. Here the two inputs, X_1 and X_2, and the output Y each take values from the set $\{0, 1\}$. Tbe conditional probability of the output Y for eacb of tbe four possible input pairs (X_1, X_2) is also sbown in tbe schema at the right in Fig. 1.

It is clear tbat if tbe transmitters can cooperate with each other they can transmit witbout error one bit per cbannel use by transmitting either tbe pair $(X_1 = 0, X_2 = 0)$ or tbe pair $(X_1 = 1, X_2 = 1)$. Sucb would be the case if a common binary source were connected to botb inputs without any coding. If a message is to be transmitted by connecting it to only one input, say to input 1, and if tbe other input is unaware of tbe message, then even if no information is to be transmitted through input 2, the information rate for input 1 must be substantially less than one bit per channel use in order to achieve reliable communication. If two independent messages are to be connected separately to the inputs—message 1 to input 1, message 2 to input 2—the situation is even more difficult.

A general configuration that we consider is sbown in Fig. 2. Three sources emitting statistically independent messages at rates R_0, R_1, and R_2 are connected to a multiple access channel via two encoders. Tbe messages from source 1 and source 0 are inputs to encoder 1 and its output is connected to one of the input terminals of the channel.

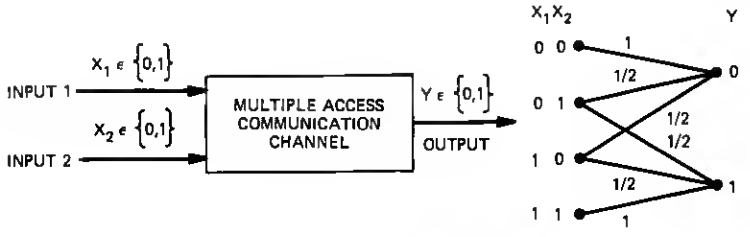

Fig. 1—A multiple access channel.

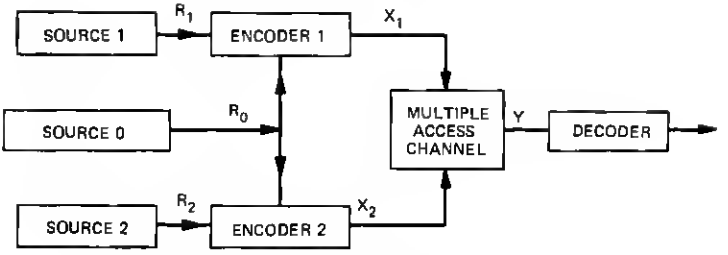

Fig. 2—Multiple access channel with correlated sources.

Encoder 2 has as inputs the messages from both source 2 and source 0 and its output is connected to the other input terminal of the channel. The channel output is connected to a decoder which estimates the three source messages. It is convenient to represent the rates of the three message sources by a point in a three-dimensional *rate space.*

For each given channel of the sort just described, there are certain rate triplets, R_0, R_1, R_2, for which it is possible to attain arbitrarily small probability of error in the system output by using sufficiently clever encoding and decoding schemes. For other points in the rate space this is not possible. We call the closure of the set of rate points for which the error probability can be made arbitrarily small the *admissible rate region* or the *capacity region* for this channel. It is a natural generalization to the multiple access channel of the channel capacity that is associated with the more commonly studied channel having a single input and a single output.

The main result of this paper is a complete determination of the capacity region $\mathcal{C}$. A typical case is shown in Fig. 3. The region always lies in the first octant and is bounded by the planes $R_0 = 0$, $R_1 = 0$,

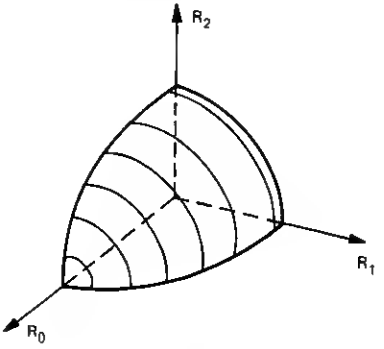

Fig. 3—An admissible rate region.

$R_2 = 0$ and a convex surface. The equations that describe $\mathcal{C}$ will be shown to involve various conditional and unconditional mutual informations. This is analogous to the single-user channel where the capacity is calculated from a mutual information.

Problems resembling ours have been treated by several authors. Shannon,[2] and then Van der Meulen,[3] consider a *two-way* channel with two inputs and two outputs. The configuration of the encoders and decoders is different than in our model, so that the problems are not the same. One similarity, however, is that the two sources are described by a pair of rates which are represented by a point in rate space. For certain points, encoders and decoders exist for which the probability of error can be made as small as desired.

The multiple access channel has been investigated by Liao,[4] Van der Meulen,[5] and Ahlswede.[6] Liao[4] and Ahlswede[6] both prove a coding theorem and a converse for the case of independent sources. Our results reduce to theirs for the case $R_0 = 0$. Correlation in the sources adds a totally new dimension to the problem (and literally to the region of admissible rates).

A problem which is the dual of the one considered here is the broadcast channel investigated by Cover[7] and Bergmans.[8] There, a channel with one input and two outputs is considered along with a single encoder and two decoders. Again the concept of an admissible rate region applies.

A brief outline of the paper follows. In Section II a detailed problem formulation is presented. Section III summarizes the main results of the paper and gives some examples. Sections IV and V and the associated appendixes give details of the derivation of a coding theorem and a converse. A more useful description of the admissible rate region is given in Section VI. We conclude in Section VII with some generalizations and comments.

II. PROBLEM FORMULATION

Consider the block diagram shown in Fig. 4. The three sources are described by a three-dimensional *rate vector* $\mathbf{R} = (R_0, R_1, R_2)$ with nonnegative components. For a fixed positive integer N, we define the components of the vector $\mathbf{M}$ by

$$\mathbf{M} = \mathbf{M}(\mathbf{R}, N) = (M_0, M_1, M_2), \tag{1a}$$

$$M_i = \lceil e^{R_i N} \rceil, \qquad i = 0, 1, 2, \tag{1b}$$

where $\lceil x \rceil$ is the smallest integer greater than or equal to x. Every N time units the sources produce a triplet of numbers (i, j, k) that are the

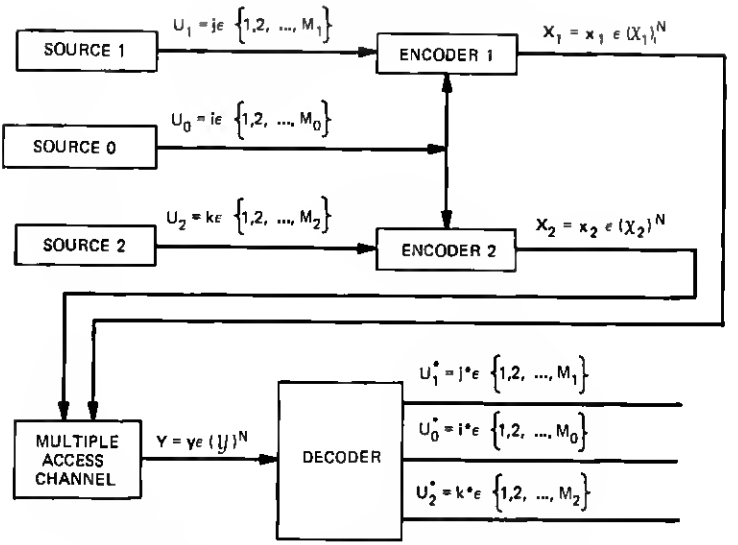

Fig. 4—Notation for multiple access channel.

corresponding values of the random variables (U_0, U_1, U_2). These random variables are assumed to be statistically independent and uniformly distributed over the rectangular lattice of dimensions $M_0 \times M_1 \times M_2$. That is, tbeir joint probability distribution is

$$\begin{aligned} P_{U_0U_1U_2}(i, j, k) = \Pr[U_0 = i, U_1 = j, U_2 = k] = 1/M_0M_1M_2, \\ i \in (1, 2, \cdots, M_0) \equiv I_0, \\ j \in (1, 2, \cdots, M_1) \equiv I_1, \\ k \in (1, 2, \cdots, M_2) \equiv I_2. \end{aligned} \tag{2}$$

The channel is a probabilistic mapping which every unit of time maps a pair of real numbers (x_1, x_2) to tbe real number y. Tbe real numbers x_1, x_2, and y belong to tbe finite alphabets $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{Y}$, respectively. Tbe mapping is governed by tbe conditional probability distribution $P_{Y|X_1X_2}(y|x_1, x_2)$ for all x_1 in $\mathcal{X}_1$, x_2 in $\mathcal{X}_2$, and y in $\mathcal{Y}$. Here we describe the inputs by the pair of random variables (X_1, X_2) and the output by the random variable Y. Throughout this paper, it will be assumed that $P_{Y|X_1X_2}$ is specified *a priori* and cannot be altered.

To describe how tbe channel processes sequences of N input pairs, we define tbe N-vectors

$$\begin{aligned} \mathbf{x}_1 &= (x_{11}, x_{12}, \cdots, x_{1N}), & \mathbf{x}_1 \in (\mathcal{X}_1)^N, \\ \mathbf{x}_2 &= (x_{21}, x_{22}, \cdots, x_{2N}), & \mathbf{x}_2 \in (\mathcal{X}_2)^N, \\ \mathbf{y} &= (y_1, y_2, \cdots, y_N), & \mathbf{y} \in (\mathcal{Y})^N, \end{aligned} \tag{3}$$

and in a similar way the corresponding random vectors $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{Y}$. Here $(\mathfrak{X}_1)^N$ is the set of all N-vectors whose components are in $\mathfrak{X}_1$. The sets $(\mathfrak{X}_2)^N$ and $(\mathfrak{Y})^N$ are defined analogously. We assume the channel is stationary and memoryless; that is,

$$P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) = \prod_{t=1}^{N} P_{Y|X_1X_2}(y_t|x_{1t}, x_{2t}). \tag{4}$$

The superscript N on the joint probability distribution indicates the dimension of the vectors.

The encoders are deterministic mappings from the source outputs to channel input vectors. Encoder 1 is a mapping from the source pair (i, j) to an N-vector $\mathbf{x}_1 \in (\mathfrak{X}_1)^N$. The functional form for this mapping is written

$$\mathbf{x}_1 = \mathbf{f}_N(i, j), \quad i \in I_0, \quad j \in I_1, \quad \mathbf{x}_1 \in (\mathfrak{X}_1)^N. \tag{5}$$

Similarly, encoder 2 is a mapping from the source pair (i, k) to the N-vector $\mathbf{x}_2 \in (\mathfrak{X}_2)^N$. The functional form for this mapping is written

$$\mathbf{x}_2 = \mathbf{g}_N(i, k), \quad i \in I_0, \quad k \in I_2, \quad \mathbf{x}_2 \in (\mathfrak{X}_2)^N. \tag{6}$$

The collection of $(M_0 \times M_1 + M_0 \times M_2)$ N-vectors which result from these mappings is called a *code* of block length N. Usually, we will adopt the more suggestive notation $\mathbf{x}_{1ij}$ and $\mathbf{x}_{2ik}$ instead of $\mathbf{f}_N(i, j)$ and $\mathbf{g}_N(i, k)$.

To summarize the operation of the sources, encoders, and channel we note that:

(*i*) Every N time units, the three sources produce a triplet (i, j, k).

(*ii*) The two encoders act upon the source outputs to produce the two N-vectors $\mathbf{x}_{1ij}$ and $\mathbf{x}_{2ik}$.

(*iii*) The components of these vectors are impressed upon the channel, one pair of inputs each time unit. Corresponding to each pair of inputs the channel produces an output, so that in the N time units the channel produces an output N-vector, $\mathbf{y}$.

The *decoder* is a deterministic mapping from the vector $\mathbf{y}$ to the triplet (i^*, j^*, k^*) where $i^* \in I_0$, $j^* \in I_1$, $k^* \in I_2$. We describe this mapping by $(i^*, j^*, k^*) = \mathbf{h}_N(\mathbf{y})$. The triplet of decoder outputs is denoted by the vector random variable (U_0^*, U_1^*, U_2^*).

The deterministic mappings $(\mathbf{f}_N, \mathbf{g}_N, \mathbf{h}_N)$ will be called a *coding*. A coding with rate vector $\mathbf{R} = (R_0, R_1, R_2)$ and block length N will be denoted by $C_N(\mathbf{R})$. For a given coding, we can in principle calculate the probability of the error event $\mathcal{E}$, where $\mathcal{E}^c$ (the complement of $\mathcal{E}$)

is defined as

$$\mathcal{E}^c = \text{event } \{U_0^* = U_0 \quad \text{and} \quad U_1^* = U_1 \quad \text{and} \quad U_2^* = U_2\}. \tag{7}$$

For the coding $C_N(\mathbf{R})$, we denote the probability of the error event (hereafter called the probability of error) by $P_e(C_N(\mathbf{R}))$.

A rate vector $\mathbf{R}$ will be said to be *admissible* if, for every $\epsilon > 0$, there exists a positive integer N and a coding $C_N(\mathbf{R})$ such that $P_e(C_N(\mathbf{R})) \leqq \epsilon$. The closure of the set of admissible rate vectors is called the *admissible region* or *capacity region*, and is denoted by $\mathcal{C}$. Our purpose is to specify $\mathcal{C}$ for an arbitrary, discrete, memoryless, multiple access channel.

III. SUMMARY OF RESULTS AND EXAMPLES

The main results of this paper are two alternative descriptions of the admissible rate region $\mathcal{C}$ for any discrete memoryless channel. The proofs that these yield the correct region are contained in the remaining sections of the paper. Here we discuss only the simplest of the results.

We shall have much need of conditional mutual information expressions in the sequel. We remind the reader of the definition

$$I(\mathbf{A}; \mathbf{B} \mid \mathbf{C}) = \sum_{\mathbf{i}} \sum_{\mathbf{j}} \sum_{\mathbf{k}} P_{\mathbf{ABC}}(\mathbf{i}, \mathbf{j}, \mathbf{k}) \log \frac{P_{\mathbf{AB}|\mathbf{C}}(\mathbf{i}, \mathbf{j} \mid \mathbf{k})}{P_{\mathbf{A}|\mathbf{C}}(\mathbf{i} \mid \mathbf{k}) P_{\mathbf{B}|\mathbf{C}}(\mathbf{j} \mid \mathbf{k})}. \tag{8}$$

Here

$$P_{\mathbf{ABC}}(\mathbf{i}, \mathbf{j}, \mathbf{k}) = \Pr [A_\alpha = i_\alpha, B_\beta = j_\beta, C_\gamma = k_\gamma, \quad \alpha = 1, 2, \cdots, L; \beta = 1, 2, \cdots, M; \gamma = 1, 2, \cdots, N]$$

is the joint distribution function of the discrete random variables $A_1, A_2, \cdots, A_L, B_1, B_2, \cdots, B_M, C_1, C_2, \cdots, C_N$. The conditional distributions $P_{\mathbf{AB}|\mathbf{C}}(\mathbf{i}, \mathbf{j} \mid \mathbf{k})$, etc., are defined in the usual way.

Let us return now to consider a discrete memoryless channel with input alphabets $\mathfrak{X}_1$ and $\mathfrak{X}_2$, output alphabet $\mathcal{Y}$, and transition probabilities $P_{Y|X_1X_2}(y \mid x_1, x_2)$, $x_1 \in \mathfrak{X}_1$, $x_2 \in \mathfrak{X}_2$, $y \in \mathcal{Y}$. Let Z be a random variable which takes on values in the set

$$\mathfrak{z} = \{1, 2, \cdots, M\}. \tag{9}$$

From any set of three distributions $P_{X_1|Z}(x_1 \mid z)$, $P_{X_2|Z}(x_2 \mid z)$, and $P_Z(z)$, $x_1 \in \mathfrak{X}_1$, $x_2 \in \mathfrak{X}_2$, $z \in \mathfrak{z}$, form the joint distribution

$$P_{ZX_1X_2Y}(z, x_1, x_2, y) = P_Z(z) P_{X_1|Z}(x_1 \mid z) P_{X_2|Z}(x_2 \mid z) P_{Y|X_1X_2}(y \mid x_1, x_2). \tag{10}$$

Now denote by $\mathfrak{R}(P_{ZX_1X_2Y})$ the set of vectors $\mathbf{R} = (R_0, R_1, R_2)$ such that

$$0 \leqq R_1 \leqq I(X_1; Y \mid X_2, Z), \tag{11a}$$

$$0 \leqq R_2 \leqq I(X_2; Y|X_1, Z), \tag{11b}$$

$$0 \leqq R_1 + R_2 \leqq I(X_1, X_2; Y|Z), \tag{11c}$$

$$0 \leqq R_0 + R_1 + R_2 \leqq I(X_1, X_2; Y), \tag{11d}$$

where the mutual informations are computed according to (8) using the joint distribution (10). This region is a polyhedron such as is shown in Fig. 7, Appendix 1. Then the admissible rate region $\mathcal{C}$ is given by

$$\mathcal{C} = \text{closure of the convex hull } \bigcup \mathcal{R}(P_{ZX_1X_2Y}), \tag{12}$$

where the union is taken over all possible choices of $P_{X_1|Z}$, $P_{X_2|Z}$, and P_Z, and all values of M, the size of the $\mathcal{Z}$ alphabet.†

To obtain the intersection of the admissible rate region $\mathcal{C}$ with the plane $R_0 = 0$, the size of the alphabet $\mathcal{Z}$ can be set equal to 1. The random variable Z no longer appears in the equations. For $R_0 = 0$, we then define $\mathcal{R}(P_{X_1}, P_{X_2})$ as the set of vector $\mathbf{R} = (0, R_1, R_2)$ such that

$$0 \leqq R_1 \leqq I(X_1; Y|X_2), \tag{13a}$$

$$0 \leqq R_2 \leqq I(X_2; Y|X_1), \tag{13b}$$

$$0 \leqq R_1 + R_2 \leqq I(X_1, X_2; Y). \tag{13c}$$

Then

$$\mathcal{C}|_{R_0=0} = \text{closure of the convex hull of } \bigcup \mathcal{R}(P_{X_1}, P_{X_2}), \tag{14}$$

where the union is taken over all possible choices for the unconditional distrihutions P_{X_1} and P_{X_2}. This is the solution found by Liao[4] for uncorrelated sources.

Other equations for specifying the region $\mathcal{C}$ are given in Section VI. They involve the calculation of mutual informations among long sequences of random variables and thus do not appear to be useful for computation.

Quite generally, $\mathcal{C}$ is convex. It is always hounded hy portions of the three coordinate planes and a surface which encloses a finite volume in the first quadrant of rate space. If $\mathbf{R} = (R_0, R_1, R_2)$ is in $\mathcal{C}$, then for any $\boldsymbol{\delta} = (\delta_0, \delta_1, \delta_2)$ satisfying $0 \leqq \delta_i \leqq R_i$, $i = 0, 1, 2$, the rate vector $\boldsymbol{\delta}$ is also in $\mathcal{C}$.

In the remainder of this section, some simple examples are presented for which the admissible rate region has an explicit characterization.

† We suspect that it suffices to consider only values of $M \leqq \lceil e^{R_0} \rceil$, but have not been able to prove this conjecture.

Example 1 (Multiplier Channel)

Both the inputs, X_1 and X_2, and the output, Y, for this channel take values 0 and 1. The output is the product of the two inputs. Formally, $\mathfrak{X}_1 = \mathfrak{X}_2 = \mathfrak{Y} = \{0, 1\}$ and $P_{Y|X_1X_2}(0|0, 0) = P_{Y|X_1X_2}(0|0, 1) = P_{Y|X_1X_2}(0|1, 0) = P_{Y|X_1X_2}(1|1, 1) = 1$, and all other conditional probabilities are zero. Note that the channel is deterministic.

The pyramid described by the planes

$$\begin{gathered} R_0 = 0, \quad R_1 = 0, \quad R_2 = 0, \\ R_0 + R_1 + R_2 = \log 2 \end{gathered} \tag{15}$$

must contain the admissible rate region $\mathcal{C}$, as is seen from (11d) since $0 \leqq R_0 + R_1 + R_2 \leqq I(X_1, X_2; Y) \leqq \log 2$. But the rate vectors $\mathbf{R}_1 = (\log 2, 0, 0)$, $\mathbf{R}_2 = (0, \log 2, 0)$, and $\mathbf{R}_3 = (0, 0, \log 2)$ are all admissible by the following strategies:

$\mathbf{R}_1$: Choose $N = 1$, $M_0 = 2$, $M_1 = M_2 = 1$ and use code words $x_{111} = 0$, $x_{121} = 1$, $x_{211} = 0$, $x_{221} = 1$.

$\mathbf{R}_2$: Choose $N = 1$, $M_0 = 1$, $M_1 = 2$, $M_2 = 1$ and use code words $x_{111} = 0$, $x_{112} = 1$, $x_{211} = 1$.

$\mathbf{R}_3$: Choose $N = 1$, $M_0 = 1$, $M_1 = 1$, $M_2 = 2$ and use code words $x_{111} = 1$, $x_{211} = 0$, $x_{212} = 1$.

The probability of error for these codes is zero. Since the convex hull of these three rate points and the origin is the set bounded by the planes (15), the capacity region $\mathcal{C}$ is as shown in Fig. 5.

By similar arguments, we find that Fig. 5 gives the region of admissible rates for many other binary-input, binary-output deterministic

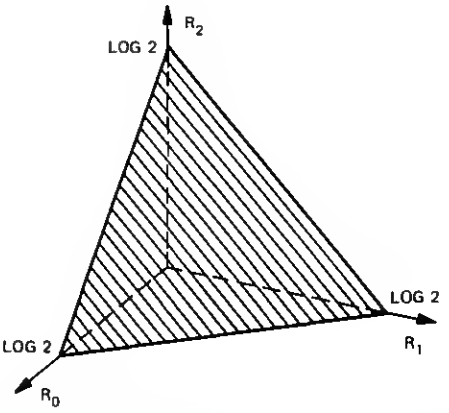

Fig. 5—Admissible rate region for the multiplier channel.

TABLE I—$P_{Y|X_1X_2}(y|x_1, x_2)$

x_1x_2 \ y	0	1	2	3
0 0	$1-p$	$p/3$	$p/3$	$p/3$
0 1	$p/3$	$1-p$	$p/3$	$p/3$
1 0	$p/3$	$p/3$	$1-p$	$p/3$
1 1	$p/3$	$p/3$	$p/3$	$1-p$

channels (ones with all transition probabilities equal to zero or one). Degenerate cases exist, however, in which the region $\mathcal{C}$ reduces to a portion of a plane. For example, if $P_{Y|X_1X_2}(0|0, 0) = P_{Y|X_1X_2}(0|0, 1) = P_{Y|X_1X_2}(1|1, 0) = P_{Y|X_1X_2}(1|1, 1) = 1$ and all other probabilities are zero, it is easy to verify that $\mathcal{C} = \{\mathbf{R} = (R_0, R_1, R_2): 0 \leqq R_0 + R_1 \leqq \log 2, R_2 = 0, R_1, R_0 \geqq 0\}$.

Example 2 (Symmetric Noisy Channel)

Let $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$, $\mathcal{Y} = \{0, 1, 2, 3\}$ and let $P_{Y|X_1X_2}(y|x_1, x_2)$ be given as shown in Table I. Let M be as in (9). Define $P_Z(z_i) = \gamma_i$, $P_{X_1|Z}(0|z_i) = \alpha_i$, $P_{X_2|Z}(0|z_i) = \beta_i$, $i = 1, 2, \cdots, M$. Straightforward calculations then yield

$$I(X_1; Y|X_2, Z) = \sum_{i=1}^{M} \gamma_i(f_1(\alpha_i, p) - K(p)), \tag{16}$$

$$I(X_2; Y|X_1, Z) = \sum_{i=1}^{M} \gamma_i(f_1(\beta_i, p) - K(p)), \tag{17}$$

$$I(X_1, X_2; Y|Z) = \sum_{i=1}^{M} \gamma_i(f_2(\alpha_i, \beta_i, p) - K(p)), \tag{18}$$

where

$$f_1(\delta, p) = \frac{2}{3}p \log \frac{3}{p} + \left((1-p)\delta + \frac{p}{3}(1-\delta)\right) \times \log \frac{1}{(1-p)\delta + \frac{p}{3}(1-\delta)} + \left(\frac{p}{3}\delta + (1-p)(1-\delta)\right) \times \log \frac{1}{\frac{p}{3}\delta + (1-p)(1-\delta)}, \tag{19}$$

$$K(p) = (1-p) \log \frac{1}{(1-p)} + p \log \frac{3}{p}, \tag{20}$$

and

$$f_2(\alpha, \beta, p) = \left[(1-p)\alpha\beta + \frac{p}{3}(1-\alpha\beta)\right] \log \frac{1}{(1-p)\alpha\beta + \frac{p}{3}(1-\alpha\beta)}$$
$$+ \left((1-p)\alpha(1-\beta) + \frac{p}{3}(1-\alpha+\alpha\beta)\right)$$
$$\times \log \frac{1}{(1-p)\alpha(1-\beta) + \frac{p}{3}(1-\alpha+\alpha\beta)}$$
$$+ \left((1-p)\beta(1-\alpha) + \frac{p}{3}(1-\beta+\alpha\beta)\right)$$
$$\times \log \frac{1}{(1-p)\beta(1-\alpha) + \frac{p}{3}(1-\beta+\alpha\beta)}$$
$$+ \left((1-p)(1-\alpha)(1-\beta) + \frac{p}{3}(\alpha+\beta-\alpha\beta)\right)$$
$$\times \log \frac{1}{(1-p)(1-\alpha)(1-\beta) + \frac{p}{3}(\alpha+\beta-\alpha\beta)}. \quad (21)$$

It is easy to show that $f_1(\delta, p) \leqq f_1(\frac{1}{2}, p)$, $f_2(\alpha, \beta, p) \leqq f_2(\frac{1}{2}, \frac{1}{2}, p) = \log 4$. Therefore, the three mutual informations in (16), (17), and (18) are simultaneously maximized by setting $\alpha_i = \beta_i = \frac{1}{2}$, $i = 1, 2, \cdots, M$. Furthermore,

$$I(X_1, X_2; Y) = H(Y) - H(Y|X_1X_2) = H(Y) - K(p) \leqq \log 4 - K(p)$$

with equality when $\alpha_i = \beta_i = \frac{1}{2}$, $i = 1, 2, \cdots, M$. Thus all four mutual informations, $I(X_1; Y|X_2, Z)$, $I(X_2; Y|X_1, Z)$, $I(X_1, X_2; Y|Z)$, and $I(X_1, X_2; Y)$, are maximized for the same choice of the parameters α_i, β_i, and γ_i, and the maximum values are independent of M. The capacity region for this channel then is given by

$$0 \leqq R_1 \leqq f_1(\tfrac{1}{2}, p) - K(p), \quad (22)$$
$$0 \leqq R_2 \leqq f_1(\tfrac{1}{2}, p) - K(p), \quad (23)$$
$$0 \leqq R_0 + R_1 + R_2 \leqq \log 4 - K(p). \quad (24)$$

This region is shown in Fig. 6.

IV. EXISTENCE OF CODINGS WITH SMALL P_e

In this section we outline a proof of the existence of codings which have vanishingly small probability of error for certain values of the

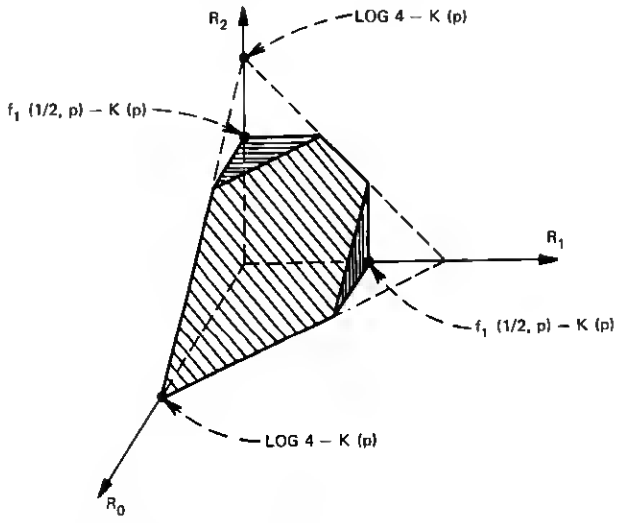

Fig. 6—Admissible rate region for the symmetric noisy binary channel.

rate vector $\mathbf{R}$ and sufficiently large block length N. Tedious details are relegated to the appendixes. A random coding argument is used. We calculate the average probability of error for an ensemble of codings, then argue that there must exist at least one memher of the ensemhle having error probability as small as this average. Actually, we can only compute an upper bound for this average error probability, but this bound is sufficiently small for our purposes.

For every coding, we shall use the same form of decoder mapping. Assume for the moment that the block length N, the rate vector $\mathbf{R}$, and the encoder functions $\mathbf{f}_N(i, j) = \mathbf{x}_{1ij}$ and $\mathbf{g}_N(i, k) = \mathbf{x}_{2ik}$ are fixed. For each $\mathbf{y} \in (\mathcal{Y})^N$, the decoder computes the $M_0 \times M_1 \times M_2$ numhers

$$P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}\,|\,\mathbf{x}_{1ij}, \mathbf{x}_{2ik}), \qquad i \in I_0, \qquad j \in I_1, \qquad k \in I_2.$$

Then $\mathbf{h}(\mathbf{y}) = (i_o, j_o, k_o)$ if and only if (i_o, j_o, k_o) is the smallest triplet (in lexicographic order) such that

$$P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}\,|\,\mathbf{x}_{1i_oj_o}, \mathbf{x}_{2i_ok_o}) \geqq P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}\,|\,\mathbf{x}_{1ij}, \mathbf{x}_{2ik}) \tag{25}$$

for all (i, j, k). Such a decoder mapping achieves a maximum likelihood decision among the possible source outputs.

We now describe the class of codings for which we obtain an upper bound to the average probability of error. It is specified by two positive integers, K and $N = KL$, where L is a positive integer, by a rate vector

$\mathbf{R}$, and by a particular probability distribution for the random variables $\mathbf{X}_1$, $\mathbf{X}_2$, and Z. The vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ are K-dimensional and take on values in $(\mathfrak{X}_1)^K$ and $(\mathfrak{X}_2)^K$ respectively, the spaces of channel input K-vectors. The random variable Z takes values from an alphabet $\mathfrak{z}$ of size M as in (9). The joint distribution of these quantities is restricted to have the form

$$P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}(z, \mathbf{x}_1, \mathbf{x}_2) = P^{(K)}_{\mathbf{X}_1|Z}(\mathbf{x}_1|z)P^{(K)}_{\mathbf{X}_2|Z}(\mathbf{x}_2|z)P_Z(z) \tag{26}$$

and we denote this collection of distributions by $\mathcal{P}_K$. A class of codings is thus specified by K, $N = KL$, $\mathbf{R}$, and a $P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2} \in \mathcal{P}_K$.

Now let K, $N = KL$, $\mathbf{R}$, and $P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2} \in \mathcal{P}_K$ be given. A set of N-dimensional code vectors $\mathbf{x}_{111}, \cdots, \mathbf{x}_{11M_1}, \mathbf{x}_{211}, \cdots, \mathbf{x}_{21M_2}$ in the corresponding ensemble of codings is obtained as follows. Choose a sample, say z, from the distribution $P_Z(\cdot)$. Next independently choose M_1 K-vectors from $P^{(K)}_{\mathbf{X}_1|Z}(\cdot\,|z)$ and then M_2 K-vectors from $P^{(K)}_{\mathbf{X}_2|Z}(\cdot\,|z)$. These are respectively the first K components of the N-vectors $\mathbf{x}_{111}, \cdots, \mathbf{x}_{11M_1}$, $\mathbf{x}_{211}, \cdots, \mathbf{x}_{21M_2}$.

To obtain the next K components of the code words, independently choose a new sample z from $P_Z(\cdot)$ and repeat the process. After a total of L drawings from $P_Z(\cdot)$ the specification of the N-vectors $\mathbf{x}_{111}, \cdots, \mathbf{x}_{11M_1}, \mathbf{x}_{211}, \cdots, \mathbf{x}_{21M_2}$ is complete. The entire process is then repeated to obtain the remaining code words—those with second subscript equal to 2, 3, $\cdots$, M_0.

We now seek an upper bound to the average probability of error for the codings in this ensemble, an average in which the probability of error for each particular coding is weighted in accordance with its probability of occurrence in the ensemble. We denote this average probability of error by $P_e(N, P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2})$ and we denote by $P_{e|i,j,k}(N, P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2})$ the average probability of error given that the source triplet (i, j, k) was presented for transmission. A useful result is

Theorem 1: The average probability of error conditioned on the source triplet (i, j, k) has an upper bound

$$P_{e|i,j,k}(N, P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}) \leqq \sum_{\alpha=1}^{4} \exp\{-N[E_\alpha(\rho_\alpha, P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}) - \rho_\alpha \hat{R}_\alpha]\}, \tag{27}$$

where $0 \leqq \rho_\alpha \leqq 1$, $\alpha = 1, 2, 3, 4$,

$$\hat{R}_\alpha = \begin{cases} R_1, & \alpha = 1 \\ R_2, & \alpha = 2 \\ R_1 + R_2, & \alpha = 3 \\ R_0 + R_1 + R_2, & \alpha = 4, \end{cases} \tag{28}$$

and

$$E_1(\rho_1, P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}) = -\frac{1}{K}\ln \sum_{\mathbf{y}}\sum_{\mathbf{x}_2}\sum_{z} P^{(K)}_{\mathbf{X}_2|Z}(\mathbf{x}_2|z)P_Z(z) \times \Big(\sum_{\mathbf{x}_1} P^{(K)}_{\mathbf{X}_1|Z}(\mathbf{x}_1|z)\big(P^{(K)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_1\mathbf{x}_2)\big)^{1/(1+\rho_1)}\Big)^{1+\rho_1}, \quad (29a)$$

$$E_2(\rho_2, P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}) = -\frac{1}{K}\ln \sum_{\mathbf{y}}\sum_{\mathbf{x}_1}\sum_{z} P^{(K)}_{\mathbf{X}_1|Z}(\mathbf{x}_1|z)P_Z(z) \times \Big(\sum_{\mathbf{x}_2} P^{(K)}_{\mathbf{X}_2|Z}(\mathbf{x}_2|z)\big(P^{(K)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)\big)^{1/(1+\rho_2)}\Big)^{1+\rho_2}, \quad (29b)$$

$$E_3(\rho_3, P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}) = -\frac{1}{K}\ln \sum_{\mathbf{y}}\sum_{z} P_Z(z) \times \Big(\sum_{\mathbf{x}_1}\sum_{\mathbf{x}_2} P^{(K)}_{\mathbf{X}_1|Z}(\mathbf{x}_1|z)P^{(K)}_{\mathbf{X}_2|Z}(\mathbf{x}_2|z) \times \big(P^{(K)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_1\mathbf{x}_2)\big)^{1/(1+\rho_3)}\Big)^{1+\rho_3}, \quad (29c)$$

$$E_4(\rho_4, P^{(K)}_{\mathbf{X}_1\mathbf{X}_2}) = -\frac{1}{K}\ln \sum_{\mathbf{y}}\Big(\sum_{\mathbf{x}_1}\sum_{\mathbf{x}_2} P^{(K)}_{\mathbf{X}_1\mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}_2) \times \big(P^{(K)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)\big)^{1/(1+\rho_4)}\Big)^{1+\rho_4} - \frac{e^{-N(R_1+R_2)}\rho_4}{N}. \quad (29d)$$

A proof of this theorem is given in Appendix A. It follows closely the proof given in Gallager[9] for tbe single-input, single-output channel.

Since tbe bound proved in tbe theorem is independent of tbe triplet (i, j, k), we see that this same bound applies to the unconditioned average probability of error $P_e(N, P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2})$. Finally, for fixed $N = KL$, and $P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}$, there must be at least one coding in the ensemble with probability of error no greater than tbe average probability of error. Thus we have

Theorem 2: For every positive integer K, for every positive integer N that is an integral multiple of K, for every joint distribution $P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}$ of form (26), and for every rate vector $\mathbf{R}$, there exists a coding $C_N(\mathbf{R})$ such that

$$P_e(C_N(\mathbf{R})) \leqq \sum_{\alpha=1}^{4} \exp\{-N[E_\alpha(\rho_\alpha, P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}) - \rho_\alpha \hat{R}_\alpha]\} \quad (30)$$

for all ρ_α, $0 \leqq \rho_\alpha \leqq 1$, $\alpha = 1, 2, 3, 4$. The E_α and $\hat{R}_\alpha$ are given by (28) and (29).

For a given $P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}$ and for certain values of tbe rate vector $\mathbf{R}$, the upper bound decreases exponentially in N. For tbese values of $\mathbf{R}$, by making N sufficiently large, we can insure a small probability of error. We now determine for what rate vectors this is the case.

Define $\mathfrak{R}(P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2\mathbf{Y}})$ as the set of rate vectors $\mathbf{R}$ for which

$$0 \leqq R_1 < \frac{1}{K} I(\mathbf{X}_1; \mathbf{Y} | \mathbf{X}_2, Z) \tag{31a}$$

$$0 \leqq R_2 < \frac{1}{K} I(\mathbf{X}_2; \mathbf{Y} | \mathbf{X}_1, Z) \tag{31b}$$

$$0 \leqq R_1 + R_2 < \frac{1}{K} I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y} | Z) \tag{31c}$$

$$0 \leqq R_0 + R_1 + R_2 < \frac{1}{K} I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}), \tag{31d}$$

where the mutual informations are evaluated under the distribution

$$P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2\mathbf{Y}}(z, \mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}(z, \mathbf{x}_1, \mathbf{x}_2) P^{(K)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y} | \mathbf{x}_1, \mathbf{x}_2), \tag{32}$$

where $P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}$ is given by (26) and $P^{(K)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}$ by (4).

In Appendix B we prove

Theorem 3: For every $\epsilon > 0$ and every rate vector $\mathbf{R} \subset \mathfrak{R}(P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2\mathbf{Y}})$, there exists an L_0 and a sequence of codings, $C_N(\mathbf{R})$, such that

$$P_e(C_N(\mathbf{R})) \leqq \epsilon \qquad \text{for every} \qquad N = KL, L \geqq L_0. \tag{33}$$

This theorem holds for all $P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}$ of the form given in (26), that is, for all $P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2} \in \mathcal{P}_K$. Now define

$$\mathfrak{R}_K \equiv \bigcup_{P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2} \in \mathcal{P}_K} \mathfrak{R}(P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2\mathbf{Y}}), \tag{34}$$

and finally define

$$\mathfrak{R} \equiv \bigcup_K \mathfrak{R}_K, \tag{35}$$

where $K = 1, 2, \cdots$. We then have the following main result:

Theorem 4: For every $\epsilon > 0$ and for every rate vector $\mathbf{R} \subset \mathfrak{R}$, there exist values of K and L_0 and a sequence of codings $C_N(\mathbf{R})$ such that

$$P_e(\mathbf{B}_N(\mathbf{R})) \leqq \epsilon \qquad \text{for every} \qquad N = KL, L \geqq L_0. \tag{36}$$

Note that if we use the statement of Theorem 1 instead of Theorem 2, Theorem 4 becomes

Corollary 1: For every $\epsilon > 0$, for every message triplet (i, j, k), and for every rate vector $\mathbf{R} \subset \mathfrak{R}$, there exist values of K and L_0 and a sequence of codings $C_N(\mathbf{R})$ such that

$$P_{e|ijk}(C_N(\mathbf{R})) \leqq \epsilon \qquad \text{for every} \qquad N = KL, L \geqq L_0. \tag{37}$$

V. CONVERSE THEOREM

In this section, we present a series of lemmas and theorems which yield a converse to the Coding Theorem 4. Let $\hat{\mathfrak{R}}$ denote the closure of the region $\mathfrak{R}$ given by (35) and let $\hat{\mathfrak{R}}^c$ be the complement of $\hat{\mathfrak{R}}$ with respect to the first octant. We shall ultimately show that every coding $C_K(\mathbf{R})$ with $\mathbf{R} \in \hat{\mathfrak{R}}^c$ transmits with a probability of error not less than a constant $\delta > 0$ which is independent of K.

Our notation is as before except that K, instead of N, will be used for the block length of a code. Let K, $\mathbf{R}$, and the channel be given. The associated vector $\mathbf{M}$ with components

$$M_\alpha = \lceil e^{KR_\alpha} \rceil, \qquad \alpha = 0, 1, 2, \tag{38}$$

is then determined. We shall no longer be concerned with ensembles of codes, but rather fix our attention on some given encoding functions $\mathbf{x}_{1ij} = \mathbf{f}(i, j)$, $\mathbf{x}_{2ik} = \mathbf{g}(i, k)$, where $i \in I_0$, $j \in I_1$, and $k \in I_2$. These vectors need not be distinct. Then with the source statistics given by (2), the given encoding defines a joint probability distribution

$$\begin{aligned} P^{(K)}_{U_0U_1U_2\mathbf{X}_1\mathbf{X}_2\mathbf{Y}}(i, j, k, \mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) &= P^{(K)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \\ &\times Q^{(K)}_{\mathbf{X}_1|U_0U_1}(\mathbf{x}_1|i, j)Q^{(K)}_{\mathbf{X}_2|U_0U_2}(\mathbf{x}_2|i, k)P_{U_0U_1U_2}(i, j, k) \end{aligned} \tag{39}$$

for the random variables in question. Here

$$Q^{(K)}_{\mathbf{X}_1|U_0U_1}(\mathbf{x}_1|i, j) = \delta_{\mathbf{x}_1\mathbf{x}_{1ij}}, \tag{40a}$$

$$Q^{(K)}_{\mathbf{X}_2|U_0U_2}(\mathbf{x}_2|i, k) = \delta_{\mathbf{x}_2\mathbf{x}_{2ik}}, \tag{40b}$$

where the right-hand terms are Kronecker deltas. Entropies and mutual informations can then be calculated from (39) by the usual formulas.

Several more definitions are needed. We shall make use of the *rate number vector* $\mathbf{R}' = (R_0', R_1', R_2')$ given by

$$R_\alpha' = \frac{1}{K} \log M_\alpha \geqq R_\alpha \qquad \alpha = 0, 1, 2 \tag{41}$$

and the elementary entropy function

$$h(x) \equiv -x \log x - (1 - x) \log (1 - x). \tag{42}$$

Finally, we define

$$P_{e1}(C_K(\mathbf{R})) = \Pr [U_1^* \neq U_1] \tag{43}$$

$$P_{e2}(C_K(\mathbf{R})) = \Pr [U_2^* \neq U_2] \tag{44}$$

$$P_{e3}(C_K(\mathbf{R})) = \Pr [U_1^* \neq U_1 \text{ or } U_2^* \neq U_2]. \tag{45}$$

Then, for the probability of error using the coding $C_K(\mathbf{R})$ we have

$$P_e(C_K(\mathbf{R})) = \Pr\left[U_0^* \neq U_0 \text{ or } U_1^* \neq U_1 \text{ or } U_2^* \neq U_2\right] \geqq \max\left[P_{e1}(C_K), P_{e2}(C_K), P_{e3}(C_K)\right]. \quad (46)$$

We now proceed to the first of the lemmas which is a generalization of Fano's inequality (Ref. 9, Theorem 4.3.1). The proof is given in Appendix C.

Lemma 1: For every K and $\mathbf{R}$ and for every $C_K(\mathbf{R})$:

$$H(U_1|\mathbf{Y}, U_0, U_2) \leqq P_{e1}(C_K) \log M_1 + h(P_{e1}(C_K)); \quad (47a)$$

$$H(U_2|\mathbf{Y}, U_0, U_1) \leqq P_{e2}(C_K) \log M_2 + h(P_{e2}(C_K)); \quad (47b)$$

$$H(U_1, U_2|\mathbf{Y}, U_0) \leqq P_{e3}(C_K) \log (M_1M_2) + h(P_{e3}(C_K)); \quad (47c)$$

$$H(U_0, U_1, U_2|\mathbf{Y}) \leqq P_e(C_K) \log (M_0M_1M_2) + h(P_e(C_K)). \quad (47d)$$

The next lemma, proved in Appendix D, is a generalization of the data processing theorem (Ref. 9, Theorem 4.3.3).

Lemma 2: For every K and $\mathbf{R}$ and every coding $C_K(\mathbf{R})$:

$$\text{(a)}, \quad I(U_1; \mathbf{Y}|U_2, U_0) \leqq I(\mathbf{X}_1; \mathbf{Y}|\mathbf{X}_2, U_0); \quad (48a)$$

$$\text{(b)}, \quad I(U_2; \mathbf{Y}|U_1, U_0) \leqq I(\mathbf{X}_2; \mathbf{Y}|\mathbf{X}_1, U_0); \quad (48b)$$

$$\text{(c)}, \quad I(U_1, U_2; \mathbf{Y}|U_0) \leqq I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}|U_0); \quad (48c)$$

$$\text{(d)}, \quad I(U_0, U_1, U_2; \mathbf{Y}) \leqq I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}). \quad (48d)$$

Lemma 3: For every K and $\mathbf{R}$ and every coding $C_K(\mathbf{R})$ with rate-number vector $\mathbf{R}'$:

$$\text{(a)}, \quad KR_1' - I(\mathbf{X}_1; \mathbf{Y}|\mathbf{X}_2, U_0) \leqq P_{e1}(C_K)KR_1' + h(P_{e1}(C_K)); \quad (49a)$$

$$\text{(b)}, \quad KR_2' - I(\mathbf{X}_2; \mathbf{Y}|\mathbf{X}_1, U_0) \leqq P_{e2}(C_K)KR_2' + h(P_{e2}(C_K)); \quad (49b)$$

$$\text{(c)}, \quad K(R_1' + R_2') - I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}|U_0) \leqq P_{e3}(C_K)K(R_1' + R_2') + h(P_{e3}(C_K)); \quad (49c)$$

$$\text{(d)}, \quad K(R_0' + R_1' + R_2') - I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}) \leqq P_e(C_K)K(R_0' + R_1' + R_2') + h(P_e(C_K)). \quad (49d)$$

The proof is given in Appendix E.

With these lemmas in hand, we return to the matter of establishing a converse to Theorem 4. For a given K, $\mathbf{R}$, and encoding $C_K(\mathbf{R})$, there is established a joint probability distribution between the random variables $\mathbf{Y}$, $\mathbf{X}_1$, $\mathbf{X}_2$, and U_0 given by

$$Q^{(K)}_{U_0\mathbf{X}_1\mathbf{X}_2\mathbf{Y}}(i, \mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = P^{(K)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_1\mathbf{x}_2)Q^{(K)}_{\mathbf{X}_1|U_0}(\mathbf{x}_1|i)Q^{(K)}_{\mathbf{X}_2|U_0}(\mathbf{x}_2|i)Q_{U_0}(i), \quad (50)$$

where $P^{(K)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}$ is given by (4),

$$Q_{U_0}(i) = \frac{1}{M_0}, \tag{51a}$$

$$Q^{(K)}_{\mathbf{X}_1|U_0}(\mathbf{x}_1|i) = \frac{1}{M_1}\sum_{j=1}^{M_1}\delta_{\mathbf{x}_1\mathbf{x}_{1ij}}, \tag{51b}$$

$$Q^{(K)}_{\mathbf{X}_2|U_0}(\mathbf{x}_2|i) = \frac{1}{M_2}\sum_{k=1}^{M_2}\delta_{\mathbf{x}_2\mathbf{x}_{2ik}}, \tag{51c}$$

$$i \in I_0, \qquad \mathbf{x}_1 \in (\mathfrak{X}_1)^K, \qquad \mathbf{x}_2 \in (\mathfrak{X}_2)^K.$$

Here $C_K(\mathbf{R})$ is defined by the code words $\mathbf{x}_{1ij}$ and $\mathbf{x}_{2ik}$, $i \in I_0$, $j \in I_1$, $k \in I_2$. We denote by $\mathcal{Q}_K$ tbe set of all distributions $Q^{(K)}_{U_0\mathbf{X}_1\mathbf{X}_2}$ derived from code books, tbat is, all distributions of tbe form obtained by summing (50) over all $\mathbf{y} \in (\mathcal{Y})^K$.

With K, $\mathbf{R}$, and an encoding $C_K(\mathbf{R})$ now fixed, we define $\mathcal{S}^c(Q^{(K)}_{U_0\mathbf{X}_1\mathbf{X}_2\mathbf{Y}})$ to be the set of all vectors $\mathbf{S} = (S_0, S_1, S_2)$ witb non-negative components such tbat *at least one* of tbe following inequalities is satisfied,

$$S_1 > \frac{1}{K} I(\mathbf{X}_1; \mathbf{Y}|U_0, \mathbf{X}_2), \tag{52a}$$

$$S_2 > \frac{1}{K} I(\mathbf{X}_2; \mathbf{Y}|U_0, \mathbf{X}_1), \tag{52b}$$

$$S_1 + S_2 > \frac{1}{K} I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}|U_0), \tag{52c}$$

$$S_0 + S_1 + S_2 > \frac{1}{K} I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}). \tag{52d}$$

Next define

$$\mathcal{S}^c_K \equiv \bigcap_{Q^{(K)}_{U_0\mathbf{X}_1\mathbf{X}_2} \in \mathcal{Q}_K} \mathcal{S}^c(Q^{(K)}_{U_0\mathbf{X}_1\mathbf{X}_2\mathbf{Y}}) \tag{53}$$

and finally

$$\mathcal{S}^c \equiv \bigcap_K \mathcal{S}^c_K. \tag{54}$$

Here c denotes complement witb respect to tbe first octant $S_0 \geqq 0$, $S_1 \geqq 0$, $S_2 \geqq 0$. Tbus, for example, $\mathcal{S}(Q^{(K)}_{U_0\mathbf{X}_1\mathbf{X}_2\mathbf{Y}})$ is a closed convex polybedron bounded by seven planes.

Note the similarity between (50) and (32) and between (31) whicb defines $\mathcal{R}(P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2\mathbf{Y}})$ and (52) whicb defines $\mathcal{S}^c(Q^{(K)}_{U_0\mathbf{X}_1\mathbf{X}_2\mathbf{Y}})$. For every distribution in $\mathcal{Q}_K$, tbere is a distribution in $\mathcal{P}_K$ tbat will give equality between the corresponding rigbt-hand members of (31) and (52) wben Z and U_0 are properly identified. We make this identification by

choosing $M = M_0$, and by taking Z to be uniformly distributed over its M possible values. With a member of $\mathcal{P}_K$ identified with each $Q^{(K)}_{U_0\mathbf{X}_1\mathbf{X}_2}$ in this way, we see that for this particular $P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}$ one has

$$\mathcal{S}(Q^{(K)}_{U_0\mathbf{X}_1\mathbf{X}_2\mathbf{Y}}) = \hat{\mathcal{R}}(P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2\mathbf{Y}}).$$

Here the caret, $\hat{}$, denotes closure. Comparison of (34) and (53) then shows that $\mathcal{S}_K \subset \hat{\mathcal{R}}_K$. Thus $\mathcal{S} \subset \hat{\mathcal{R}}$ or

$$\hat{R}^c \subset \mathcal{S}^c. \tag{55}$$

In Appendix F we establish

Theorem 5: If $\mathbf{R}$ *is an interior point of* $\mathcal{S}^c$, *then for every* K *and every encoding* $C_K(\mathbf{R})$,

$$P_e(C_K(\mathbf{R})) \geqq \delta > 0,$$

where $\delta = \delta(\mathbf{R})$ *is independent of the encoding and of* K.

VI. SPECIFICATION OF THE CAPACITY REGION

At the end of Section II, the capacity region was defined as the closure of the set of admissible rate points. Theorems 4 and 5 along with (55) show that $\mathcal{C} = \hat{\mathcal{R}}$ where $\mathcal{R}$ is defined by (31), (34), and (35). This characterization of $\mathcal{C}$ is of little computational value. It entails the calculation of the mutual informations appearing on the right of (31) for all distributions of form (32). A further infinite union over all values of K is then required. In this section we shall show how a much simpler description of $\mathcal{C}$ can be obtained, one that is independent of K and hence much more suitable for numerical calculations.

Central to the development of this simpler characterization of $\mathcal{C}$ is

Theorem 6: The region $\mathcal{C}$ *of admissible rates is convex.*

This theorem is proved by a time-sharing argument in Appendix G. By deleting words from a code, one obtains an additional obvious feature of the region $\mathcal{C}$ which we state as

Theorem 7: Let $\mathbf{R} \in \mathcal{C}$. *Then if* $0 \leqq R''_\alpha \leqq R_\alpha$, $\alpha = 0, 1, 2$, *the rate vector* $\mathbf{R}''$ *is also contained in* $\mathcal{C}$.

We return to our simpler characterization of $\mathcal{C}$. Let $\mathcal{R}_1$ denote the region specified by (31), (32), and (34) for $K = 1$. Since $\mathcal{C}$ is the closure of $\mathcal{R}$ as given by (35), $\hat{\mathcal{R}}_1 \subseteq \mathcal{C}$. From Theorem 6 it follows that also

$$\hat{\mathcal{R}}' = \text{convex hull } \mathcal{R}_1 \subseteq \mathcal{C}. \tag{56}$$

(The convex hull of a set $\mathcal{A}$ consists of all points in $\mathcal{A}$ and all points

on all line segments joining points of $\mathcal{C}$.) We shall soon show that indeed $\mathcal{R}' = \mathcal{C}$.

As a step in this direction, in Appendix H we establish

Lemma 4: For every distribution $P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2\mathbf{Y}}(z, \mathbf{x}_1, \mathbf{x}_2, \mathbf{y})$ *as in* (32),

$$I(\mathbf{X}_1; \mathbf{Y}|Z, \mathbf{X}_2) \leqq \sum_{t=1}^{K} I(X_{1t}; Y_t|Z, X_{2t}), \tag{57a}$$

$$I(\mathbf{X}_2; \mathbf{Y}|Z, \mathbf{X}_1) \leqq \sum_{t=1}^{K} I(X_{2t}; Y_t|Z, X_{1t}), \tag{57b}$$

$$I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}|Z) \leqq \sum_{t=1}^{K} I(X_{1t}, X_{2t}; Y_t|Z), \tag{57c}$$

$$I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}) \leqq \sum_{t=1}^{K} I(X_{1t}, X_{2t}; Y_t). \tag{57d}$$

Combined with (31) the lemma shows that

$$\mathcal{R}^*(P^{(K)}_{Z\mathbf{X}_1|\mathbf{X}_2\mathbf{Y}}) \supseteq \mathcal{R}(P^{(K)}_{Z\mathbf{X}_1|\mathbf{X}_2\mathbf{Y}}), \tag{58}$$

where $\mathcal{R}^*(P^{(K)}_{Z\mathbf{X}_1|\mathbf{X}_2\mathbf{Y}})$ is the set of rate vectors $\mathbf{R}$ for which

$$0 \leqq R_1 \leqq \frac{1}{K}\sum_{1}^{K} I(X_{1t}; Y_t|Z, X_{2t}), \tag{59a}$$

$$0 \leqq R_2 \leqq \frac{1}{K}\sum_{1}^{K} I(X_{2t}; Y_t|Z, X_{1t}), \tag{59b}$$

$$0 \leqq R_1 + R_2 \leqq \frac{1}{K}\sum_{1}^{K} I(X_{1t}, X_{2t}; Y_t|Z), \tag{59c}$$

$$0 \leqq R_0 + R_1 + R_2 \leqq \frac{1}{K}\sum_{1}^{K} I(X_{1t}, X_{2t}; Y_t), \tag{59d}$$

where the right sides of (59) are evaluated under $P^{(K)}_{Z\mathbf{X}_1|\mathbf{X}_2\mathbf{Y}}$. Note that $\mathcal{R}^*(P^{(K)}_{Z\mathbf{X}_1|\mathbf{X}_2\mathbf{Y}})$, unlike $\mathcal{R}(P^{(K)}_{Z\mathbf{X}_1|\mathbf{X}_2\mathbf{Y}})$, is a closed set by definition.

Now, a typical term on the right of (59) depends only on the marginal distribution $P_{ZX_{1t}X_{2t}Y_t}(z, x_{1t}, x_{2t}, y_t)$. By summing (32) over the appropriate indexes and taking account of (26), it is seen that this marginal can be written

$$P_{ZX_{1t}X_{2t}Y_t}(z, x_{1t}, x_{2t}, y_t) = P_Z(z)P_{X_{1t}|Z}(x_{1t}|z)P_{X_{2t}|Z}(x_{2t}|z)P_{Y_t|X_{1t}X_{2t}}(y_t|x_{1t}, x_{2t})$$

which is a distribution of the form $P^{(K)}_{Z\mathbf{X}_1|\mathbf{X}_2\mathbf{Y}}$ for $K = 1$. Thus the right-hand sides of (59), which are the parameters defining the box-like

region $\mathfrak{R}^*(P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2Y})$, are averages of parameters that define the box-like region $\mathfrak{R}(P^{(1)}_{ZX_{1t}X_{2t}Y_t})$, $t = 1, 2, \cdots, K$. In Appendix I this fact and the convexity of the box-like regions are used to show that

$$\text{convex hull} \bigcup_{t=1}^{K} \hat{\mathfrak{R}}(P^{(1)}_{ZX_{1t}X_{2t}Y_t}) \supseteq \mathfrak{R}^*(P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2Y}) \tag{60}$$

from which it also follows that

$$\text{convex hull} \bigcup_{P^{(1)}_{ZX_1X_2} \in \mathcal{P}_1} \hat{\mathfrak{R}}(P^{(1)}_{ZX_1X_2Y}) \supseteq \bigcup_{P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2} \in \mathcal{P}_K} \mathfrak{R}^*(P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2Y_1}). \tag{61}$$

We now have

$$\begin{aligned}\mathfrak{R}' \equiv \text{convex hull } \hat{\mathfrak{R}}_1 &= \text{convex hull closure} \bigcup_{P^{(1)}_{ZX_1X_2} \in \mathcal{P}_1} \mathfrak{R}(P^{(1)}_{ZX_1X_2Y}) \\ &\supseteq \text{convex hull} \bigcup_{P^{(1)}_{ZX_1X_2} \in \mathcal{P}_1} \hat{\mathfrak{R}}(P^{(1)}_{ZX_1X_2Y}) \supseteq \bigcup_{P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2} \in \mathcal{P}_K} \mathfrak{R}^*(P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2Y}) \\ &\supseteq \bigcup_{P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2} \in \mathcal{P}_K} \mathfrak{R}(P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2Y}). \end{aligned} \tag{62}$$

Here the last inclusion follows from (58) and the next to last inclusion is (61). Using (34) and (62), we now see that $\mathfrak{R}' \supseteq \mathfrak{R}_K$ for every K. From (35) then $\mathfrak{R}' \supseteq \mathfrak{R}$, and since $\mathfrak{R}'$ is closed by definition, $\mathfrak{R}' \supseteq \hat{\mathfrak{R}} = \mathcal{C}$. Combined with (56), this shows that $\mathfrak{R}' = \mathcal{C}$, and the formulation (10)–(12) is thereby established.

It is to he noted that while this reduction permits calculation of $\mathcal{C}$ by evaluating mutual informations involving no more than four random variables, the size of the Z alphabet is unrestricted. In this connection, see the footnote in Section III.

That we can indeed take the size of the Z alphahet to he 1 when computing the intersection of $\mathcal{C}$ with the plane $R_0 = 0$, as claimed in Section III, is seen as follows. When $R_0 = 0$, (11d) is weaker than (11c), since always $I(X_1, X_2; Y) \geqq I(X_1, X_2; Y|Z)$. Thus we need only consider (11a), (11b), and (11c) in defining regions $\mathfrak{R}(P_{ZX_1X_2Y})$ in the $R_1 - R_2$ plane. But the right members of these equations are of the form

$$\begin{aligned} I(X_1; Y|X_2, Z) &= \sum_i P_Z(z_i) I(X_1; Y|X_2, Z = z_i) \\ I(X_2; Y|X_1, Z) &= \sum_i P_Z(z_i) I(X_2; Y|X_1, Z = z_i) \\ I(X_1, X_2; Y|Z) &= \sum_i P_Z(z_i) I(X_1, X_2; Y|Z = z_i). \end{aligned}$$

An argument just like that of Appendix I now shows that $\mathfrak{R} \subseteq$ convex hull $\bigcup_i \mathfrak{R}_i$ where $\mathfrak{R}_i$ is given by $0 \leqq R_1 \leqq I(X_1; Y|X_2, Z = z_i)$,

$0 \leqq R_2 \leqq I(X_2; Y|X_1, Z = z_i)$, $0 \leqq R_1 + R_2 \leqq I(X_1, X_2; Y|Z = z_i)$. Each box-like region $\mathfrak{R}_i$ can be thought of as obtained from a distribution in which Z takes a single value with probability one. The formulation of Section III follows at once.

VII. COMMENTARY

7.1 *Generalizations*

7.1.1 *N Input Users*

The foregoing can be generalized to the case of a memoryless channel with N input users and a single output. The channel then is specified by alphabets $\mathcal{Y}$, $\mathcal{X}_1$, $\mathcal{X}_2$, $\cdots$, $\mathcal{X}_N$ and transition probabilities $p(y|x_1, x_2, \cdots, x_N)$ for $y \in \mathcal{Y}$, $x_i \in \mathcal{X}_i$, $i = 1, 2, \cdots, N$. Again we allow the information supplied to the input users to be correlated in a special way.

We first write out the equations for $N = 3$ in full, and then indicate the general result. There are now seven independent sources, S_1, S_2, S_3, S_{12}, S_{13}, S_{23}, S_{123} producing information at rates R_1, R_2, R_3, R_{12}, R_{13}, R_{23}, R_{123} respectively. There are three encoders. Encoder 1 sees the outputs of only S_1, S_{12}, S_{13}, S_{123}; encoder 2 sees the outputs of only S_2, S_{12}, S_{23}, S_{123}; encoder 3 sees the outputs of only S_3, S_{13}, S_{23}, S_{123}. The decoder at the channel output attempts to reproduce separately the messages from the seven sources. Using block codes, for certain values of the rate vector $\mathbf{R} = (R_1, R_2, R_3, R_{12}, R_{13}, R_{23}, R_{123})$, the error probability of the system can be made arbitrarily small. The closure of the set of all such vector rates is called the *capacity region* $\mathfrak{C}$.

$\mathfrak{C}$ can be found as follows. Let

$$p_{123}(z_{123}),\ p_{12}(z_{12}),\ p_{13}(z_{13}),\ p_{23}(z_{23})$$
$$p_1(x_1|z_{123}, z_{12}, z_{13}),\ p_2(x_2|z_{123}, z_{12}, z_{23}),\ p_3(x_3|z_{123}, z_{13}, z_{23}) \tag{63}$$

be given probability distributions. Here $x_i \in \mathcal{X}_i$, $i = 1, 2, 3$. The Z_{12}, Z_{13}, etc., have finite alphabets of unspecified size. We denote by P the distribution

$$P = p_{123}(z_{123})p_{12}(z_{12})p_{13}(z_{13})p_{23}(z_{23})p_1(x_1|z_{123}, z_{12}, z_{13}) \times p_2(x_2|z_{123}, z_{12}, z_{23})p_3(x_3|z_{123}, z_{13}, z_{23})p(y|x_1, x_2, x_3). \tag{64}$$

Now let $\mathfrak{R}(P)$ be the set of $\mathbf{R}$ such that

$$0 \leqq R_1 \leqq I(X_1; Y|Z_{123}, Z_{12}, Z_{13}, Z_{23}, X_2, X_3)$$
$$0 \leqq R_2 \leqq I(X_2; Y|Z_{123}, Z_{12}, Z_{13}, Z_{23}, X_1, X_3)$$
$$0 \leqq R_3 \leqq I(X_3; Y|Z_{123}, Z_{12}, Z_{13}, Z_{23}, X_1, X_2)$$

$$0 \leqq R_1 + R_2 \leqq I(X_1, X_2; Y | Z_{123}, Z_{12}, Z_{13}, Z_{23}, X_3)$$
$$0 \leqq R_1 + R_3 \leqq I(X_1, X_3; Y | Z_{123}, Z_{12}, Z_{13}, Z_{23}, X_2)$$
$$0 \leqq R_2 + R_3 \leqq I(X_2, X_3; Y | Z_{123}, Z_{12}, Z_{13}, Z_{23}, X_1)$$
$$0 \leqq R_1 + R_2 + R_3 \leqq I(X_1, X_2, X_3; Y | Z_{123}, Z_{12}, Z_{13}, Z_{23})$$
$$0 \leqq R_1 + R_2 + R_3 + R_{12} \leqq I(X_1, X_2, X_3; Y | Z_{123}, Z_{13}, Z_{23})$$
$$0 \leqq R_1 + R_2 + R_3 + R_{13} \leqq I(X_1, X_2, X_3; Y | Z_{123}, Z_{12}, Z_{23})$$
$$0 \leqq R_1 + R_2 + R_3 + R_{23} \leqq I(X_1, X_2, X_3; Y | Z_{123}, Z_{12}, Z_{13})$$
$$0 \leqq R_1 + R_2 + R_3 + R_{12} + R_{13} \leqq I(X_1, X_2, X_3; Y | Z_{123}, Z_{23})$$
$$0 \leqq R_1 + R_2 + R_3 + R_{12} + R_{23} \leqq I(X_1, X_2, X_3; Y | Z_{123}, Z_{13})$$
$$0 \leqq R_1 + R_2 + R_3 + R_{13} + R_{23} \leqq I(X_1, X_2, X_3; Y | Z_{123}, Z_{12})$$
$$0 \leqq R_1 + R_2 + R_3 + R_{12} + R_{13} + R_{23} \leqq I(X_1, X_2, X_3; Y | Z_{123})$$
$$0 \leqq R_1 + R_2 + R_3 + R_{12} + R_{13} + R_{23} + R_{123} \leqq I(X_1, X_2, X_3; Y), \quad (65)$$

where all the mutual informations here are computed with the distribution (64). Let $\mathfrak{R} = \bigcup \mathfrak{R}(P)$ where the union is over all distributions of form (64) as the factors listed in (63) are varied. Then $\mathfrak{C}$ is the closure of the convex bull of $\mathfrak{R}$.

Tbe generalization to N users is simple in concept but awkward to describe. We do not dwell long on it here. Tbere are now $2^N - 1$ sources and $\mathfrak{C}$ is a region in a $(2^N - 1)$-dimensional rate space. Tbe list (63) is increased to contain $2^N - N - 1$ separate distributions for as many independent Z variables—$Z_{12}, Z_{13}, \cdots, Z_{23}, \cdots, Z_{123\cdots N}$—and N distributions of form $p_1(x_1 | z_{12}, z_{13}, \cdots, z_{12\cdots N})$, etc., where each z subscript contains the x subscript. Equations (64) and (65) are generalized in an obvious way. There are now $\sum_{j=1}^{N} \left(2^{\binom{N}{j}} - 1 \right)$ equations (65). $\mathfrak{C}$ is given as the closure of the convex hull of the union of the regions defined by tbese equations.

Tbese results for N users were obtained by cursory examination of tbe rigorous proofs given in this paper for two users. As we have not had the courage to write out all the details, however, the assertions made for the N-user case must still be regarded as conjectures, or educated guesses.

7.1.2 *Continuous Amplitudes*

It would appear that our results can be extended in a natural way to channels with more general alphabet structures. For example, the cbannel might be specified by a conditional probability *density* $P(y | x_1, x_2)$ wbere x_1, x_2, and y take all real values. Equation (11)

would remain the same, but the mutual informations are now given by integrals. Densities $P_Z(z)$, $P_{X_1|Z}(x_1|z)$, $P_{X_2|Z}(x_2|z)$ must be specified and the joint density of Z, X_1, X_2, and Y is the product (10) as before. Constraints, such as $EX_1^2 = \sigma_1^2$, $EX_2^2 = \sigma_2^2$ must be imposed on these densities in taking the union indicated in (12).

Again, we have not verified in detail the validity of the determination of $\mathfrak{C}$ just given for continuous amplitudes. *Caveat emptor.*

7.2 *Some Problems*

Many research problems related to the subject of this paper remain to be examined. A brief description of some of these follows:

(*i*) The footnote in Section III suggests that the size of the alphabet Z can be bounded in searching for the capacity of a particular channel. Is this conjecture true?

(*ii*) The explicit construction of good codes for use on specific multiple access channels is an untapped field that leads to new problems not found on single-input, single-output channels. For example, even for noiseless channels (all channel probabilities zero or one) a coding problem exists since users compete with each other for the use of the channel.

(*iii*) The region of rates for which error-free transmission with finite length codes is possible is not known. This region is analogous to the zero-error capacity of the single-input, single-output channel.

(*iv*) For a particular multiple access channel it has been found that the region of admissible rates can be enlarged by allowing the encoders to observe the output via a feedback channel. This is in contrast to the situation for the single-input, single-output channel where feedback does not alter the capacity. In the multi-user case, however, a feedback channel increases the cooperation possible between the users and in general increases the forward capacity. How to calculate the region of admissible rates for multiple access channels with feedback is not known.

(*v*) A special form has been assumed here for the correlation between the messages encoded by the two users. How does one handle more general correlations? Is the presently assumed form general in some asymptotic sense?

(*vi*) Can one calculate the capacity region for some class of multiple access channels with memory?

(*vii*) What is the rate distortion theory for these channels?

VIII. ACKNOWLEDGMENTS

We are deeply indebted to Professor N. T. Gaarder of the University of Hawaii for the many fruitful discussions we bad with him on all aspects of this research. We also wish to acknowledge the work of Henry Liao whose investigation of the case of uncorrelated sources led us to the present study.

APPENDIX A

Proof of Theorem 1

Let $P_{e|i,j,k}(C)$ be tbe probability of error when the source triplet (i, j, k) is sent over tbe cbannel using coding C. Let Pr (C) be tbe probability of the particular coding C. Tben

$$P_{e|i,j,k}(N, P^{(K)}_{\mathbf{X}_1\mathbf{X}_2}) = \sum \Pr(C) P_{e|i,j,k}(C), \tag{66}$$

where the sum is over all possible codings, that is, over all ways of choosing the code words $\mathbf{x}_{111}, \cdots, \mathbf{x}_{1M_0M_1}, \mathbf{x}_{211}, \cdots, \mathbf{x}_{2M_0M_2}$. But tbe right side of (66) can be interpreted as the probability of error in tbe joint experiment of drawing a code from the ensemble and transmitting (i, j, k) over tbe cbannel. Witb this interpretation in mind, we have

$$P_{e|i,j,k}(N, P^{(K)}_{\mathbf{X}_1\mathbf{X}_2}) = \sum_{i=1}^{4} P_i, \tag{67}$$

where

$$P_1 = \Pr[U_0^* = i, U_1^* \neq j, U_2^* = k | \mathfrak{B}] \tag{68a}$$

$$P_2 = \Pr[U_0^* = i, U_1^* = j, U_2^* \neq k | \mathfrak{B}] \tag{68b}$$

$$P_3 = \Pr[U_0^* = i, U_1^* \neq j, U_2^* \neq k | \mathfrak{B}] \tag{68c}$$

$$P_4 = \Pr[U_0^* \neq i | \mathfrak{B}], \tag{68d}$$

where $\mathfrak{B}$ is the event $\{U_0 = i, U_1 = j, U_2 = k\}$. We will find upper bounds for tbese four probabilities.

We first compute an upper bound for P_1. Fix values for the N-vectors $\mathbf{y}, \mathbf{x}_{1ij}, \mathbf{x}_{2ik}$. Let $\mathbf{z}_i$ denote tbe L-vector whose components $z_{i1}, z_{i2}, \cdots, z_{iL}$ were used in the choice of $\mathbf{x}_{1ij}$ and $\mathbf{x}_{2ik}$. Later we shall average over these quantities.

Define $\mathfrak{A}_{j'}$ as tbe event that

$$P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_1}(\mathbf{y} | \mathbf{X}_{1ij'}, \mathbf{x}_{2ik}) \geqq P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y} | \mathbf{x}_{1ij}, \mathbf{x}_{2ik}). \tag{69}$$

Note that the only random variable in this expression is $\mathbf{X}_{1ij'}$. Define

$$P^{(N)(K)}_{\mathbf{X}|Z}(\mathbf{x}|\mathbf{z}) = \prod_{\alpha=1}^{L} P^{(K)}_{\mathbf{X}|Z}(\mathbf{x}_\alpha | z_\alpha), \tag{70}$$

where $\mathbf{x}$ is the N-vector obtained by concatenating the L K-dimensional vectors $\mathbf{x}_\alpha$, and $\mathbf{z}$ is an L-vector whose components are z_α, $\alpha = 1, 2, \cdots, L$. Then the probability of the event $\mathcal{A}_{j'}$ is

$$\Pr[\mathcal{A}_{j'}] = \sum_{\mathbf{x}_{1ij'}}' P^{(N,K)}_{\mathbf{X}_1|Z}(\mathbf{x}_{1ij'}|\mathbf{z}_i),$$

where the sum is over all values of $\mathbf{x}_{1ij'}$ satisfying

$$P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_{1ij'}, \mathbf{x}_{2ik}) \geqq P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_{1ij}, \mathbf{x}_{2ik}). \tag{71}$$

Following Gallager,[9] an upper bound to this expression is

$$\Pr[\mathcal{A}_{j'}] \leqq \sum_{\mathbf{x}_{1ij'}} P^{(N,K)}_{\mathbf{X}_1|Z}(\mathbf{x}_{1ij'}|\mathbf{z}_i) \left(\frac{P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_{1ij'}, \mathbf{x}_{2ik})}{P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_{1ij}, \mathbf{x}_{2ik})} \right)^{s_1}, \tag{72}$$

for any $s_1 \geqq 0$. The summation is over all values of the N-vector $\mathbf{x}_{1ij'}$.

For the same fixed values of $\mathbf{y}$, $\mathbf{x}_{1ij}$, $\mathbf{x}_{2ik}$, and $\mathbf{z}_i$, let $\mathcal{A}$ he the event that (69) holds for *some* value of j' not equal to j. Then from Gallager[9] (page 136)

$$\Pr[\mathcal{A}] \leqq \left(\sum_{\substack{j'=1 \\ j' \neq j}}^{M_1} \Pr[\mathcal{A}_{j'}] \right)^{\rho_1}, \tag{73}$$

for any ρ_1 in the range $0 \leqq \rho_1 \leqq 1$. Combining (72) and (73) we have

$$\Pr[\mathcal{A}] \leqq (M_1 - 1)^{\rho_1} \left[\sum_{\mathbf{x}_1} P^{(N,K)}_{\mathbf{X}_1|Z}(\mathbf{x}_1|\mathbf{z}_i) \times \left(\frac{P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_{2ik})}{P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_{1ij}, \mathbf{x}_{2ik})} \right)^{s_1} \right]^{\rho_1}, \tag{74}$$

where the summation is over all N-vectors in $(\mathcal{X}_1)^N$.

The probability of interest, P_1, has an upper hound

$$P_1 \leqq \sum_{\mathbf{y}} \sum_{\mathbf{x}_{1ij}} \sum_{\mathbf{x}_{2ik}} \sum_{\mathbf{z}_i} P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_{1ij}, \mathbf{x}_{2ik}) P^{(N,K)}_{\mathbf{X}_1|Z}(\mathbf{x}_{1ij}|\mathbf{z}_i) \times P^{(N,K)}_{\mathbf{X}_2|Z}(\mathbf{x}_{2ik}|\mathbf{z}_i) P^{(L)}_{Z}(\mathbf{z}_i) \Pr[\mathcal{A}], \tag{75}$$

where the inequality results from the fact that the occurrence of the event $\mathcal{A}$ does not necessarily imply the event $\{U_0^* = i, U_1^* \neq j, U_2^* = k\}$ but that the converse is true. Combining (74) and (75) and choosing $s_1 = 1/(1 + \rho_1)$, we obtain

$$P_1 \leqq (M_1 - 1)^{\rho_1} \sum_{\mathbf{y}} \sum_{\mathbf{x}_2} \sum_{\mathbf{z}} P^{(N,K)}_{\mathbf{X}_2|Z}(\mathbf{x}_2|\mathbf{z}) P^{(L)}_{Z}(\mathbf{z}) \times \left[\sum_{\mathbf{x}_1} (P^{(N,K)}_{\mathbf{X}_1|Z}(\mathbf{x}_1|\mathbf{z}) P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2))^{1/(1+\rho_1)} \right]^{1+\rho_1}, \tag{76}$$

where the summations for $\mathbf{y}$, $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{z}$ are taken over all elements in the spaces $(\mathcal{Y})^N$, $(\mathcal{X}_1)^N$, $(\mathcal{X}_2)^N$, and $(\mathcal{Z})^L$ respectively.

We note from (1h) that $(M_1 - 1) < e^{NR_1}$. Now use the product form of (70) and write the right-hand side of (76) as an exponential of a logarithm. We find the desired result

$$P_1 \leqq \exp\{-N[E_1(\rho_1, P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}) - \rho_1 R_1]\}, \tag{77}$$

where E_1 is given by (29a). The sums there are over all $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{y}$, and z contained respectively in $(\mathcal{X}_1)^K$, $(\mathcal{X}_2)^K$, $(\mathcal{Y})^K$, and $\mathcal{Z}$. Reversing the role of U_1 and U_2 one immediately obtains

$$P_2 \leqq \exp\{-N(E_2(\rho_2, P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}) - \rho_2 R_2)\}, \tag{78}$$

where E_2 is given by (29b).

The procedure for ohtaining the upper bound for P_3 is very similar to that used for P_1. An outline of the proof follows. Fix $\mathbf{y}$, $\mathbf{x}_{1ij}$, $\mathbf{x}_{2ik}$, and $\mathbf{z}_i$. Define $\mathfrak{B}$ as the event

$$P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{X}_{1ij'}, \mathbf{X}_{2ik'}) \geqq P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_{1ij}, \mathbf{x}_{2ik}) \quad \text{for some } j' \neq j \text{ and some } k' \neq k. \tag{79}$$

It can then be shown that for any $s_3 \geqq 0$ and $0 \leqq \rho_3 \leqq 1$

$$\Pr[\mathfrak{B}] \leqq (M_1 - 1)^{\rho_3}(M_2 - 1)^{\rho_3}\Bigg[\sum_{\mathbf{x}_1}\sum_{\mathbf{x}_2} P^{(N,K)}_{\mathbf{X}_1|Z}(\mathbf{x}_1|\mathbf{z}_i) \times P^{(N,K)}_{\mathbf{X}_2|Z}(\mathbf{x}_2|\mathbf{z}_i)\left(\frac{P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)}{P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_{1ij}, \mathbf{x}_{2ik})}\right)^{s_3}\Bigg]^{\rho_3}. \tag{80}$$

Averaging over $\mathbf{y}$, $\mathbf{x}_{1ij}$, $\mathbf{x}_{2ik}$, and $\mathbf{z}_i$, and then setting $s_3 = 1/(1 + \rho_3)$, we obtain

$$P_3 \leqq (M_1 - 1)^{\rho_3}(M_2 - 1)^{\rho_3}\sum_{\mathbf{y}}\sum_{\mathbf{z}} P^{(L)}_{Z}(\mathbf{z}) \times \Big[\sum_{\mathbf{x}_1}\sum_{\mathbf{x}_2} P^{(N,K)}_{\mathbf{X}_1|Z}(\mathbf{x}_1|\mathbf{z}) P^{(N,K)}_{\mathbf{X}_2|Z}(\mathbf{x}_2|\mathbf{z})(P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2))^{1/(1+\rho_3)}\Big]^{1+\rho_3}. \tag{81}$$

Replace $(M_1 - 1)^{\rho_3}(M_2 - 1)^{\rho_3}$ by the upper bound $e^{N\rho_3(R_1+R_2)}$, use (70) repeatedly, and write terms as exponentials of logarithms. One finds

$$P_3 \leqq \exp\{-N[E_3(\rho_3, P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}) - \rho_3(R_1 + R_2)]\}, \tag{82}$$

where E_3 is given by (29c).

One minor change is made in the procedure to compute the upper hound for P_4. We fix only the values of $\mathbf{y}$, $\mathbf{x}_{1ij}$, and $\mathbf{x}_{2ik}$ (but not of $\mathbf{z}_i$). Define $\mathfrak{D}$ as the event

$$P^{(N,K)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{X}_{1i'j'}, \mathbf{X}_{2i'k'}) \geqq P^{(N,K)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_{1ij}, \mathbf{x}_{2ik}) \tag{83}$$

for some $i' \neq i$ and any j' and k'. Then for any $s_4 \geqq 0$, $0 \leqq \rho_4 \leqq 1$,

$$\Pr[\mathfrak{D}] \leqq (M_0 - 1)^{\rho_4}(M_1)^{\rho_4}(M_2)^{\rho_4} \times \left[\sum_{\mathbf{x}_1}\sum_{\mathbf{x}_2} P^{(N,K)}_{\mathbf{X}_1\mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}_2)\left(\frac{P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}\,|\,\mathbf{x}_1, \mathbf{x}_2)}{P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}\,|\,\mathbf{x}_{1ij}, \mathbf{x}_{2ik})}\right)^{s_4}\right]^{\rho_4}, \quad (84)$$

where

$$P^{(N,K)}_{\mathbf{X}_1\mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}_2) = \sum_{\mathbf{z}} P^{(N,K)}_{\mathbf{Z}\mathbf{X}_1\mathbf{X}_2}(\mathbf{z}, \mathbf{x}_1, \mathbf{x}_2). \quad (85)$$

Averaging over $\mathbf{y}$, $\mathbf{x}_{1ij}$, $\mathbf{x}_{2ik}$ and setting $s_4 = 1/(1 + \rho_4)$, we obtain

$$P_4 \leqq (M_0 - 1)^{\rho_4}(M_1)^{\rho_4}(M_2)^{\rho_4} \times \sum_{\mathbf{y}}\left[\sum_{\mathbf{x}_1}\sum_{\mathbf{x}_2} P^{(N,K)}_{\mathbf{X}_1\mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}_2)(P^{(N)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}\,|\,\mathbf{x}_1, \mathbf{x}_2)^{1/(1+\rho_4)}\right]^{1+\rho_4}. \quad (86)$$

From (1b) we see that $(M_0 - 1) < e^{NR_0}$. An upper bound for M_1 follows from (1b) as

$$M_1 < e^{NR_1} + 1 = e^{NR_1}(1 + e^{-NR_1}) = \exp\left\{N\left[R_1 + \frac{\log(1 + e^{-NR_1})}{N}\right]\right\} \leqq \exp\left\{N\left[R_1 + \frac{e^{-NR_1}}{N}\right]\right\}. \quad (87)$$

Using a similar upper bound for M_2, we have that

$$(M_0 - 1)(M_1)(M_2) \leqq \exp\left\{N\left[R_0 + R_1 + R_2 + \frac{e^{-N(R_1+R_2)}}{N}\right]\right\}. \quad (88)$$

From (88) and (70) we then obtain

$$P_4 \leqq \exp\{-N[E_4(\rho_4, P^{(K)}_{\mathbf{X}_1\mathbf{X}_2}) - \rho_4(R_0 + R_1 + R_2)]\}, \quad (89)$$

where E_4 is given by (29d). Summing (77), (78), (82), and (89) results in (27) which was to be proved.

APPENDIX B

Proof of Theorem 3

It can be easily verified that

$$E_\alpha(\rho_\alpha, P^{(K)}_{\mathbf{Z}\mathbf{X}_1\mathbf{X}_2})\,|_{\rho_\alpha=0} = 0 \qquad \text{for} \quad \alpha = 1, 2, 3, 4. \quad (90)$$

It can also be shown by a straightforward but tedious calculation that

$$\left.\frac{\partial E_\alpha}{\partial \rho_\alpha}\right|_{\rho_\alpha=0} = \begin{cases} \frac{1}{K} I(\mathbf{X}_1; \mathbf{Y}\,|\,\mathbf{X}_2, Z), & \alpha = 1 \\ \frac{1}{K} I(\mathbf{X}_2; \mathbf{Y}\,|\,\mathbf{X}_1, Z), & \alpha = 2 \\ \frac{1}{K} I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}\,|\,Z), & \alpha = 3 \\ \frac{1}{K} I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}) - \frac{e^{-N(R_1+R_2)}}{N}, & \alpha = 4 \end{cases} \quad (91)$$

where the I's are mutual informations among K-vectors as computed under the joint distribution (32). Furthermore, from (29) it is seen that E_α is analytic in ρ_α in the neighborhood of $\rho_\alpha = 0$ and so can be expanded in a Taylor series about this point, $\alpha = 1, 2, 3, 4$:

$$E_\alpha(\rho_\alpha, P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2}) = \begin{cases} 0 + \frac{1}{K} I_\alpha \rho_\alpha + O_\alpha(\rho_\alpha^2), & \alpha = 1, 2, 3 \\ 0 + \left[\frac{1}{K} I_4 - \frac{e^{-N(R_1+R_2)}}{N} \right] \rho_4 + O_4(\rho_4^2), & \alpha = 4. \end{cases} \tag{92}$$

Here I_α is the appropriate expression from (91) and O is the usual Bachmann-Landau order-of-magnitude symbol. Furthermore, if $\mathbf{R} \subset \mathfrak{R}(P^{(K)}_{Z\mathbf{X}_1\mathbf{X}_2\mathbf{Y}})$, we have from (31) that

$$\frac{1}{K} I_\alpha - R_\alpha \equiv \delta_\alpha > 0, \qquad \alpha = 1, 2, 3\ 4. \tag{93}$$

Combining (92) and (93) with (30), we see that

$$P_e(C_N) \leqq \sum_{\alpha=1}^{3} \exp\{-N\rho_\alpha[\delta_\alpha + O_\alpha(\rho_\alpha^2)/\rho_\alpha]\} + \exp\left\{-N\rho_4\left[\delta_4 - \frac{e^{-N(R_1+R_2)}}{N} + O_4(\rho_4^2)/\rho_4\right]\right\}. \tag{94}$$

Now choose the integer $\hat{L}$ so large that

$$\bar{\delta}_4 \equiv \delta_4 - \frac{e^{-\hat{L}K(R_1+R_2)}}{\hat{L}K}$$

is positive. Next, choose sufficiently small positive values of $\rho_1, \rho_2, \rho_3, \rho_4$ so that $\delta_\alpha + O_\alpha(\rho_\alpha^2)/\rho_\alpha > 0$, $\alpha = 1, 2, 3$, and $\bar{\delta}_4 + O_4(\rho_4^2)/\rho_4 > 0$. The coefficient of N in each exponential of (94) is now negative, and we can increase N in multiples of K starting at $N = K\hat{L}$ until each term of (94) is less than $\epsilon/4$. Call this value of N, $N_0 = KL_0$. Then (33) follows. Q.E.D.

APPENDIX C

Proof of Lemma 1

By definition of $P_{e1}(C_K)$ and $H(U_1|\mathbf{Y}, U_0, U_2)$,

$$P_{e1}(C_K) = \sum_{\mathbf{y}} \sum_{i} \sum_{j \neq j^*(\mathbf{y})} \sum_{k} P^{(K)}_{U_0U_1U_2\mathbf{Y}}(i, j, k, \mathbf{y}), \tag{95}$$

and

$$H(U_1|\mathbf{Y}, U_0, U_2) = \sum_{\mathbf{y}} \sum_{i} \sum_{j} \sum_{k} P^{(K)}_{U_0U_1U_2\mathbf{Y}}(i, j, k, \mathbf{y}) \times \log \frac{1}{P^{(K)}_{U_1|U_0U_2\mathbf{Y}}(j|i, k, \mathbf{y})}. \quad (96)$$

By separating out the terms for which $j = j^*(\mathbf{y})$ in (96), one finds the identity

$$\begin{aligned} T &\equiv H(U_1|\mathbf{Y}, U_0, U_2) - P_{e1}(C_K) \log (M_1 - 1) - h(P_{e1}(C_K)) \\ &= \sum_{\mathbf{y}} \sum_{i} \sum_{j \neq j^*(\mathbf{y})} \sum_{k} P^{(K)}_{U_0U_1U_2\mathbf{Y}}(i, j, k, \mathbf{y}) \\ &\quad \times \log \frac{P_{e1}(C_K)}{(M_1 - 1)P^{(K)}_{U_1|U_0U_2\mathbf{Y}}(j|i, k, \mathbf{y})} \\ &\quad + \sum_{\mathbf{y}} \sum_{i} \sum_{k} P^{(K)}_{U_0U_1U_2\mathbf{Y}}(i, j^*(\mathbf{y}), k, \mathbf{y}) \\ &\quad \times \log \frac{(1 - P_{e1}(C_K))}{P^{(K)}_{U_1|U_0U_2\mathbf{Y}}(j^*(\mathbf{y})|i, k, \mathbf{y})}. \end{aligned} \quad (97)$$

Now use the fact that $\log x \leqq x - 1$ to obtain

$$\begin{aligned} T &\leqq \sum_{\mathbf{y}} \sum_{i} \sum_{j \neq j^*(\mathbf{y})} \sum_{k} \left[\frac{P_{e1}P^{(K)}_{U_0U_2\mathbf{Y}}(i, k, \mathbf{y})}{(M_1 - 1)} - P_{U_0U_1U_2\mathbf{Y}}(i, j, k, \mathbf{y}) \right] \\ &\quad + \sum_{\mathbf{y}} \sum_{i} \sum_{k} [(1 - P_{e1})P^{(K)}_{U_1|U_0U_2\mathbf{Y}}(j^*|i, k, \mathbf{y}) \\ &\quad - P_{U_0U_1U_2\mathbf{Y}}(i, j^*(\mathbf{y}), k, \mathbf{y})] \\ &= P_{e1} + (1 - P_{e1}) - 1 = 0. \end{aligned} \quad (98)$$

Replacing M_1 by $M_1 + 1$ yields (47a). Equations (47b), (47c), and (47d) are proved in a similar way starting from the definitions

$$P_{e2}(C_K) = \sum_{\mathbf{y}} \sum_{i} \sum_{j} \sum_{k \neq k^*(\mathbf{y})} P^{(K)}_{U_0U_1U_2\mathbf{Y}}(i, j, k, \mathbf{y}), \quad (99a)$$

$$P_{e3}(C_K) = \sum_{\mathbf{y}} \sum_{i} \sum_{(j,k) \neq (j^*(\mathbf{y}), k^*(\mathbf{y}))} P^{(K)}_{U_0U_1U_2\mathbf{Y}}(i, j, k, \mathbf{y}), \quad (99b)$$

and

$$P_e(C_K) = \sum_{\mathbf{y}} \sum_{(i,j,k) \neq (i^*, j^*, k^*)} P^{(K)}_{U_0U_1U_2\mathbf{Y}}(i, j, k, \mathbf{y}). \quad (99c)$$

APPENDIX D

Proof of Lemma 2

For part (a), we write a complicated conditional mutual information in two different ways:

$$\begin{aligned} &I(\mathbf{X}_1, U_1; \mathbf{Y}|U_2, \mathbf{X}_2, U_0) \\ &\quad = I(\mathbf{X}_1; \mathbf{Y}|U_2, \mathbf{X}_2, U_0) + I(U_1; \mathbf{Y}|\mathbf{X}_1, U_2, \mathbf{X}_2, U_0) \\ &\quad = I(U_1; \mathbf{Y}|U_2, \mathbf{X}_2, U_0) + I(\mathbf{X}_1; \mathbf{Y}|U_1, U_2, \mathbf{X}_2, U_0). \end{aligned} \quad (100)$$

Now

$$I(\mathbf{X}_1; \mathbf{Y} | U_0, U_2, \mathbf{X}_2)$$
$$= E\left\{\log \frac{P^{(K)}_{\mathbf{Y}|U_0U_1U_2\mathbf{X}_1\mathbf{X}_2}(\mathbf{Y}|U_0, U_1, U_2, \mathbf{X}_1, \mathbf{X}_2)}{P^{(K)}_{\mathbf{Y}|U_0U_2\mathbf{X}_2}(\mathbf{Y}|U_0, U_2, \mathbf{X}_2)}\right\}$$
$$= E\left\{\log \frac{P^{(K)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{Y}|\mathbf{X}_1, \mathbf{X}_2)}{P^{(K)}_{\mathbf{Y}|U_0\mathbf{X}_2}(\mathbf{Y}|U_0, \mathbf{X}_2)}\right\} = I(\mathbf{X}_1; \mathbf{Y}|U_0, \mathbf{X}_2), \quad (101)$$

where the equalities result from the special form of the joint distributions as given by (39) and (40). For the next mutual information in (100), we have

$$I(U_1; \mathbf{Y}|U_0, U_2, \mathbf{X}_1, \mathbf{X}_2)$$
$$= E\left\{\log \frac{P^{(K)}_{\mathbf{Y}|U_0U_1U_2\mathbf{X}_1\mathbf{X}_2}(\mathbf{Y}|U_0, U_1, U_2, \mathbf{X}_1, \mathbf{X}_2)}{P^{(K)}_{\mathbf{Y}|U_0U_2\mathbf{X}_1\mathbf{X}_2}(\mathbf{Y}|U_0, U_2, \mathbf{X}_1, \mathbf{X}_2)}\right\}$$
$$= E\left\{\log \frac{P^{(K)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{Y}|\mathbf{X}_1, \mathbf{X}_2)}{P^{(K)}_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{Y}|\mathbf{X}_1, \mathbf{X}_2)}\right\} = 0. \quad (102)$$

The third mutual information in (100) can he written

$$I(U_1; \mathbf{Y}|U_0, U_2, \mathbf{X}_2)$$
$$= E\left\{\log \frac{P^{(K)}_{\mathbf{Y}|U_0U_1U_2\mathbf{X}_2}(\mathbf{Y}|U_0, U_1, U_2, \mathbf{X}_2)}{P^{(K)}_{\mathbf{Y}|U_0U_2\mathbf{X}_2}(\mathbf{Y}|U_0, U_2, \mathbf{X}_2)}\right\}$$
$$= E\left\{\log \frac{P^{(K)}_{\mathbf{Y}|U_0U_1U_2}(\mathbf{Y}|U_0, U_1, U_2)}{P^{(K)}_{\mathbf{Y}|U_0U_2}(\mathbf{Y}|U_0, U_2)}\right\} = I(U_1; \mathbf{Y}|U_0, U_2). \quad (103)$$

Finally,

$$I(\mathbf{X}_1; \mathbf{Y}|U_0, U_1, U_2, \mathbf{X}_2) \geqq 0, \quad (104)$$

since all mutual informations are non-negative. Comhining (100)–(104), we obtain (48a) which completes the proof of part (a).

The proofs for parts (b), (c), and (d) follow in a similar manner.

The equations corresponding to (100) are:
Part (b),

$$I(U_2, \mathbf{X}_2; \mathbf{Y}|U_0, U_1, \mathbf{X}_1)$$
$$= I(\mathbf{X}_2; \mathbf{Y}|U_0, U_1, \mathbf{X}_1) + I(U_2; \mathbf{Y}|U_0, U_1, \mathbf{X}_1, \mathbf{X}_2)$$
$$= I(U_2; \mathbf{Y}|U_0, U_1, \mathbf{X}_1) + I(\mathbf{X}_2; \mathbf{Y}|U_0, U_1, U_2, \mathbf{X}_1); \quad (105)$$

Part (c),

$$I(U_1, U_2, \mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}|U_0)$$
$$= I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}|U_0) + I(U_1, U_2; \mathbf{Y}|U_0, \mathbf{X}_1, \mathbf{X}_2)$$
$$= I(U_1, U_2; \mathbf{Y}|U_0) + I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}|U_0, U_1, U_2); \quad (106)$$

Part (d),

$$I(U_0, U_1, U_2, \mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}) = I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}) + I(U_0, U_1, U_2; \mathbf{Y}|\mathbf{X}_1, \mathbf{X}_2) = I(U_0, U_1, U_2; \mathbf{Y}) + I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}|U_0, U_1, U_2). \quad (107)$$

Q.E.D.

APPENDIX E

Proof of Lemma 3

We use the identities:

$$\text{(a)}, \quad I(U_1; \mathbf{Y}|U_0, U_2) = H(U_1|U_0, U_2) - H(U_1|U_0, U_1, \mathbf{Y}); \quad (108a)$$

$$\text{(b)}, \quad I(U_2; \mathbf{Y}|U_0, U_1) = H(U_2|U_0, U_1) - H(U_2|U_0, U_1, \mathbf{Y}); \quad (108b)$$

$$\text{(c)}, \quad I(U_1, U_2; \mathbf{Y}|U_0) = H(U_1, U_2|U_0) - H(U_1, U_2|U_0, \mathbf{Y}); \quad (108c)$$

$$\text{(d)}, \quad I(U_0, U_1, U_2; \mathbf{Y}) = H(U_0, U_1, U_2) - H(U_0, U_1, U_2|\mathbf{Y}). \quad (108d)$$

From the joint distributions of the random variables U_0, U_1, and U_2 given in (2), we have:

$$\text{(a)}, \quad H(U_1|U_0, U_1) = H(U_1) = KR_1'; \quad (109a)$$

$$\text{(b)}, \quad H(U_2|U_0, U_1) = H(U_2) = KR_2'; \quad (109b)$$

$$\text{(c)}, \quad H(U_1, U_2|U_0) = H(U_1, U_2) = H(U_1) + H(U_2) = K(R_1' + R_2'); \quad (109c)$$

$$\text{(d)}, \quad H(U_0, U_1, U_2) = H(U_0) + H(U_1) + H(U_2) = K(R_0' + R_1' + R_2'). \quad (109d)$$

Combining the appropriate equations in (108), (109), (41), (47), and (48), we have (49) which was to be proved.

APPENDIX F

Proof of Theorem 5

If $\mathbf{R}$ is an interior point of $\mathcal{S}^c$, then there is a sphere, σ, of radius $\eta(\mathbf{R}) > 0$, centered on $\mathbf{R}$ such that every point in σ is also in $\mathcal{S}^c$. Thus every point in $\mathcal{S}(Q^{(K)}_{U_0\mathbf{X}_1\mathbf{X}_2\mathbf{Y}})$ must be distant more than $\eta(\mathbf{R})$ away from $\mathbf{R}$, and this is true for every K, and every $Q^{(K)}_{U_0\mathbf{X}_1\mathbf{X}_2}$ in $\mathcal{Q}_K$. This in turn

implies that one of the inequalities

$$\begin{aligned} R_1 - \frac{1}{K} I(\mathbf{X}_1; \mathbf{Y} \mid U_0, \mathbf{X}_2) &> \eta(\mathbf{R}) \\ R_2 - \frac{1}{K} I(\mathbf{X}_2; \mathbf{Y} \mid U_0, \mathbf{X}_1) &> \eta(\mathbf{R}) \\ R_1 + R_2 - \frac{1}{K} I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y} \mid U_0) &> \eta(\mathbf{R}) \\ R_0 + R_1 + R_2 - \frac{1}{K} I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}) &> \eta(\mathbf{R}) \end{aligned} \tag{110}$$

must hold for every encoding $C_K(\mathbf{R})$ of the sort under consideration, whenever $\mathbf{R}$ is interior to $\mathcal{S}^e$.

Now from (41), $R_\alpha' \geqq R_\alpha$, $\alpha = 0, 1, 2$, so that one of (110) holds also when the R's are replaced by R''s. From Lemma 3 we then find that

$$\hat{R}_\alpha' P_{e\alpha}(C_K(\mathbf{R})) + \frac{1}{K} h(P_{e\alpha}(C_K(\mathbf{R}))) > \eta(\mathbf{R}) \quad \text{for at least one } \alpha, \quad \alpha = 1, 2, 3, 4, \tag{111}$$

where we define

$$P_{e4}(C_K(\mathbf{R})) = P_e(C_K(\mathbf{R})) \tag{112}$$

and for any rate vector $\mathbf{R}$ we define an associated 4-vector $\hat{\mathbf{R}}$ by

$$(\hat{R}_1, \hat{R}_2, \hat{R}_3, \hat{R}_4) = (R_1, R_2, R_1 + R_2, R_0 + R_1 + R_2). \tag{113}$$

But from (87) and (88) we see that

$$\hat{R}_\alpha' \leqq \hat{R}_\alpha + \frac{e^{-K\hat{R}_\alpha}}{K} \leqq (\hat{R}_\alpha + e^{-\hat{R}_\alpha}), \tag{114}$$

so that

$$\hat{R}_\alpha' P_{e\alpha}(C_K(\mathbf{R})) + \frac{h(P_{e\alpha}(C_K(\mathbf{R}))}{K} \leqq (\hat{R}_\alpha + e^{-\hat{R}_\alpha}) P_{e\alpha}(C_K(\mathbf{R})) + h(P_{e\alpha}(C_K(\mathbf{R}))). \tag{115}$$

Combining (111) and (115) we find that

$$(\hat{R}_\alpha + e^{-\hat{R}_\alpha}) P_{e\alpha}(C_K(\mathbf{R})) + h(P_{e\alpha}(C_K(\mathbf{R})) \geqq \eta(\mathbf{R}) \quad \text{for at least one } \alpha, \quad \alpha = 1, 2, 3, 4. \tag{116}$$

Now

$$h(x) \leqq 2\sqrt{x}, \qquad 0 \leqq x \leqq 1, \tag{117}$$

as can he seen hy the following simple argument. From

$$0 < \tfrac{1}{2}[1 + (1 - Z)^2],$$

it follows that $2Z < 1 + Z + Z^2/2 \leqq e^Z$, for $Z \geqq 0$. But for $Z < 0$ we also clearly have $2Z < e^Z$ so that $2Z < e^Z$ for all Z. Suhstitute $Z = \log \sqrt{(1-t)/t}$ to ohtain

$$\log \frac{1-t}{t} < \frac{\sqrt{1-t}}{\sqrt{t}} < \frac{1}{\sqrt{t}}, \qquad 0 < t < 1$$

or

$$\int_\epsilon^x \log \frac{1-t}{t}\, dt < \int_\epsilon^x \frac{dt}{\sqrt{t}}, \qquad \epsilon < x < 1.$$

Perform the integration and take the limit as $\epsilon \to 0$. Equation (117) results.

Use (117) in (116) to find that $(\hat{R}_\alpha + e^{-\hat{R}_\alpha})P_{e\alpha} + 2\sqrt{P_{e\alpha}} \geqq \eta(\mathbf{R})$ for at least one α, $\alpha = 1, 2, 3, 4$. This implies that

$$P_{e\alpha}(C_K(\mathbf{R})) \geqq \left[\frac{\sqrt{1 + \eta(\mathbf{R})[\hat{R}_\alpha + e^{-\hat{R}_\alpha}]} - 1}{\hat{R}_\alpha + e^{-\hat{R}_\alpha}} \right]^2 \equiv \delta_\alpha(\mathbf{R}) > 0.$$

Since $P_e(C_K(\mathbf{R})) \geqq \max_\alpha [P_{e\alpha}(C_K(\mathbf{R}))]$, we find finally that

$$P_e(C_K(\mathbf{R})) \geqq \delta(\mathbf{R}) > 0, \tag{118}$$

where $\delta(\mathbf{R}) \equiv \min_\alpha \delta_\alpha(\mathbf{R})$ is independent of K and the encoding $C_K(\mathbf{R})$. Q.E.D.

APPENDIX G

Proof of Theorem 6

We first show that for every positive integer n and every integer r such that $0 \leqq r \leqq n$,

$$\mathbf{R}_1 \in \mathfrak{R}, \quad \mathbf{R}_2 \in \mathfrak{R} \Rightarrow \mathbf{R}_3 \equiv \frac{r}{n}\mathbf{R}_1 + \frac{n-r}{n}\mathbf{R}_2 \in \mathfrak{R}. \tag{119}$$

Since $\mathfrak{C}$ is the closure of $\mathfrak{R}$, and since the rationals are dense in the reals, (119) implies that if $\mathbf{R}_1 \in \mathfrak{C}$ and $\mathbf{R}_2 \in \mathfrak{C}$, then for every λ, $0 \leqq \lambda \leqq 1$, $\mathbf{R}_3 \equiv \lambda\mathbf{R}_1 + (1-\lambda)\mathbf{R}_2 \in \mathfrak{C}$, which shows $\mathfrak{C}$ to he convex.

To estahlish (119), we use the notion of time sharing to generate new codings from old ones. Suppose we have two codings $C_N(\mathbf{R}_1)$ and $C_N(\mathbf{R}_2)$ hoth of hlock length N and with numbers of words $\mathbf{M}_1$ and $\mathbf{M}_2$

respectively, where as usual

$$\mathbf{M}_\alpha = (M_{0\alpha}, M_{1\alpha}, M_{2\alpha}) = (\lceil e^{NR_{0\alpha}} \rceil, \lceil e^{NR_{1\alpha}} \rceil, \lceil e^{NR_{2\alpha}} \rceil) \qquad \alpha = 1, 2. \tag{120}$$

Denote by P_{e1} and P_{e2} the respective error probabilities achievable with $C_N(\mathbf{R}_1)$ and $C_N(\mathbf{R}_2)$. Now consider the possible channel input vectors that can be obtained by using $C_N(\mathbf{R}_1)$ r times followed by $(n - r)$ uses of $C_N(\mathbf{R}_2)$. The totality of these input vectors, each of nN components, can be thought of as the words of a new code of block length nN. Denoting its word size parameter by $\mathbf{M}$, we have

$$M_i = (M_{i1})^r (M_{i2})^{n-r}, \qquad i = 0, 1, 2. \tag{121}$$

If we use the decoders for $C_N(\mathbf{R}_1)$ and $C_N(\mathbf{R}_2)$ to decode the appropriate blocks of length N in this new larger code, the error probability for the new code, P_e, will satisfy

$$1 - P_e = (1 - P_{e1})^r (1 - P_{e2})^{n-r} \geqq (1 - rP_{e1})[1 - (n - r)P_{e2}] \geqq 1 - rP_{e1} - (n - r)P_{e2}$$

so that

$$P_e \leqq rP_{e1} + (n - r)P_{e2}. \tag{122}$$

Here we have used the fact that the channel is memoryless.

We now use this time-sharing notion to establish (119). Suppose that integers n and r are given with $n > 0$, $0 \leqq r \leqq n$ and that $\mathbf{R}_1$ and $\mathbf{R}_2$ are rate points in $\mathfrak{R}$. Suppose further that $\epsilon > 0$ is given. Then, from Theorem 4, there exist positive integers K_1 and L_1 and a sequence of codings $C_{N_1}(\mathbf{R}_1)$, $N_1 = K_1L_1$, $K_1(L_1 + 1)$, $K_1(L_1 + 2)$, $\cdots$ such that for each coding of the sequence $P_e(C_{N_1}(\mathbf{R}_1)) > \epsilon/n$. Similarly there exist integers K_2 and L_2 and a second sequence of codings $C_{N_2}(\mathbf{R}_2)$, $N_2 = K_2L_2$, $K_2(L_2 + 1)$, $K_2(L_2 + 2)$, $\cdots$ such that for each coding in the sequence $P_e(C_{N_2}(\mathbf{R}_2)) < \epsilon/n$. We now choose one coding out of each of these sequences of codings in such a way that they are of the same block length N. A suitable choice for N is the least common multiple of K_1L_1 and K_2L_2. Call the two codings $C_N(\mathbf{R}_1)$ and $C_N(\mathbf{R}_2)$. Their error probabilities are $P_{e1} < \epsilon/n$ and $P_{e2} < \epsilon/n$. Time sharing them as discussed earlier yields a new coding C, of block length $N_3 = nN$, code book size $\mathbf{M}$ given by (121) and (120), and error probability

$$P_e \leqq rP_{e1} + (n - r)P_{e2} = r\frac{\epsilon}{n} + (n - r)\frac{\epsilon}{n} = \epsilon$$

from (122). Now from the fact that $\lceil x \rceil \lceil y \rceil \geqq \lceil xy \rceil$, (121) and (120) give

$$M_i = \lceil e^{NR_{i1}} \rceil^r \lceil e^{NR_{i2}} \rceil^{n-r} \geqq \lceil e^{N[rR_{i1}+(n-r)R_{i2}]} \rceil \geqq \lceil e^{N_3 R_{i3}} \rceil, \qquad i = 0, 1, 2,$$

where $\mathbf{R}_3$ is as in (119). Thus by deleting some words from the code C we can obtain a coding with rate $\mathbf{R}_3$, and block length N_3, that has error probability $P_e < \epsilon$. Q.E.D.

APPENDIX H

Proof of Lemma 4

Consider (57a). We write

$$\begin{aligned} I(\mathbf{X}_1; \mathbf{Y} | Z, \mathbf{X}_2) &= E \log \frac{P^{(K)}_{\mathbf{Y}|Z\mathbf{X}_1\mathbf{X}_2}(\mathbf{Y}|Z, \mathbf{X}_1, \mathbf{X}_2)}{P^{(K)}_{\mathbf{Y}|Z\mathbf{X}_2}(\mathbf{Y}|Z, \mathbf{X}_2)} \\ &= E \log \frac{\prod_{t=1}^{K} P_{Y|X_1X_2}(Y_t|X_{1t}, X_{2t})}{\prod_{t=1}^{K} P^{(K)}_{Y_t|Z\mathbf{X}_2Y_1\cdots Y_{t-1}}(Y_t|Z, \mathbf{X}_2, Y_1, \cdots, Y_{t-1})} \\ &= \sum_{t=1}^{K} [H(Y_t|Z, \mathbf{X}_2, Y_2, \cdots, Y_{t-1}) - H(Y_t|X_{1t}, X_{2t})]. \end{aligned} \tag{123}$$

Here, for $t = 1$, the conditioning on $Y_1, \cdots, Y_{t-1}$ is to be omitted. But

$$H(Y_t|Z, \mathbf{X}_2, Y_1, \cdots Y_{t-1}) \leqq H(Y_t|Z, X_{2t}), \tag{124}$$

since removing conditioning random variables cannot decrease an entropy. Combining (123) and (124) we have

$$I(\mathbf{X}_1; \mathbf{Y}|Z, \mathbf{X}_2) \leqq \sum_{t=1}^{K} [H(Y_t|Z, X_{2t}) - H(Y_t|X_{1t}, X_{2t})], \tag{125}$$

or

$$I(\mathbf{X}_1; \mathbf{Y}|Z, \mathbf{X}_2) \leqq \sum_{t=1}^{K} I(X_{1t}; Y_t|Z, X_{2t}). \tag{126}$$

The proofs for (57b), (57c), and (57d) are similar.

APPENDIX I

Let numbers A_t, B_t, C_t, and D_t be given that satisfy the inequalities

$$0 \leqq A_t \leqq C_t, \tag{127a}$$

$$0 \leqq B_t \leqq C_t, \tag{127b}$$

$$0 \leqq C_t \leqq A_t + B_t, \tag{127c}$$

$$0 \leqq C_t \leqq D_t, \qquad t = 1, 2, \cdots, K. \tag{127d}$$

Let $\mathcal{R}_t$ denote the set of points (x, y, z) in three-space such that

$$0 \leqq x \leqq A_t, \tag{128a}$$

$$0 \leqq y \leqq B_t, \tag{128b}$$

$$0 \leqq x + y \leqq C_t, \tag{128c}$$

$$0 \leqq x + y + z \leqq D_t, \tag{128d}$$

for $t = 1, 2, \cdots, K$. A sketch of $\mathcal{R}_t$ is shown in Fig. 7, corresponding to the case in which all the inequalities in (127) are strict. We further define

$$\mathcal{R} \equiv \bigcup_{t=1}^{K} \mathcal{R}_t. \tag{129}$$

Now consider the region $\mathcal{R}_0$ consisting of all points (x, y, z) such that

$$0 \leqq x \leqq A_0 \equiv \frac{1}{K} \sum_{t=1}^{K} A_t, \tag{130a}$$

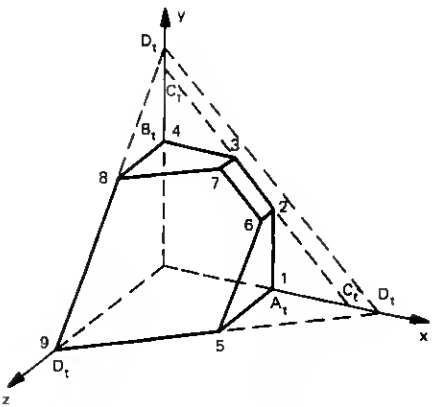

Fig. 7—The convex region $\mathcal{R}_t$.

$$0 \leqq y \leqq B_0 \equiv \frac{1}{K} \sum_{t=1}^{K} B_t, \tag{130h}$$

$$0 \leqq x + y \leqq C_0 \equiv \frac{1}{K} \sum_{t=1}^{K} C_t, \tag{130c}$$

$$0 \leqq x + y + z \leqq D_0 \equiv \frac{1}{K} \sum_{t=1}^{K} D_t. \tag{130d}$$

Our first goal in this appendix is to show that

$$\mathcal{R}_0 \subseteq \text{convex hull } \mathcal{R}. \tag{131}$$

By summing the inequalities (127) and using the definitions of A_0, B_0, C_0, and D_0 given in (130), we see that (127) also holds for $t = 0$. $\mathcal{R}_0$, too, then has the form shown in Fig. 7. As is seen, each region $\mathcal{R}_t$, $t = 0, 1, \cdots, K$, is convex and has ten extreme points, the components of which are listed below:

$$\begin{aligned}
\mathbf{r}_{0t} &= (0, 0, 0) \\
\mathbf{r}_{1t} &= (A_t, 0, 0) \\
\mathbf{r}_{2t} &= (A_t, C_t - A_t, 0) \\
\mathbf{r}_{3t} &= (C_t - B_t, B_t, 0) \\
\mathbf{r}_{4t} &= (0, B_t, 0) \\
\mathbf{r}_{5t} &= (A_t, 0, D_t - A_t) \\
\mathbf{r}_{6t} &= (A_t, C_t - A_t, D_t - C_t) \\
\mathbf{r}_{7t} &= (C_t - B_t, B_t, D_t - C_t) \\
\mathbf{r}_{8t} &= (0, B_t, D_t - B_t) \\
\mathbf{r}_{9t} &= (0, 0, D_t).
\end{aligned} \tag{132}$$

[Some of these points may coincide if there are equalities in (127) instead of strict inequalities.] For the extreme point of $\mathcal{R}_0$ we also have

$$\mathbf{r}_{i0} = \frac{1}{K} \sum_{t=1}^{K} \mathbf{r}_{it}, \qquad i = 0, 1, \cdots, 9 \tag{133}$$

which follows directly from (132) and the definitions on the right of (130). We recall that a convex body is characterized by its extreme

points: $\mathbf{r} \in \mathcal{R}_t$ if and only if

$$\mathbf{r} = \sum_{i=0}^{9} \lambda_i \mathbf{r}_{it}, \tag{134}$$

where

$$\lambda_i \geqq 0, \quad i = 0, 1, \cdots, 9 \quad \text{and} \quad \sum_{0}^{9} \lambda_i = 1, \quad t = 0, 1, \cdots, 9. \tag{135}$$

Equation (131) is now easy to establish. It is clear that the convex hull of $\mathcal{R}$ is the set of all points that can be written in the form

$$\mathbf{r}' = \sum_{i=0}^{9} \sum_{t=1}^{K} u_{it} \mathbf{r}_{it}, \tag{136}$$

where

$$u_{it} \geqq 0, \quad i = 0, 1, \cdots, 9, \quad t = 1, \cdots, K, \quad \sum_{i=0}^{9} \sum_{t=1}^{K} u_{it} = 1. \tag{137}$$

Now let $\mathbf{r}$ be any element in $\mathcal{R}_0$. Then $\mathbf{r}$ can be written in the form (134)–(135) with $t = 0$. Substituting from (133) yields

$$\mathbf{r} = \sum_{i=0}^{9} \sum_{t=1}^{K} \frac{\lambda_i}{K} \mathbf{r}_{it}. \tag{138}$$

But defining

$$u'_{it} = \frac{\lambda_i}{K}, \qquad i = 0, \cdots, 9, \qquad t = 1, 2, \cdots, K, \tag{139}$$

we see that $u'_{it} \geqq 0$ and

$$\sum_{i=0}^{9} \sum_{t=0}^{K} u'_{it} = 1. \tag{140}$$

Comparison with (136) now shows that $\mathbf{r}$ is in the convex hull of $\mathcal{R}$. Equation (131) then follows.

The application of the foregoing to (60) is immediate. Let

$$\begin{aligned} A_t &= I(X_{1t}; Y_t | X_{2t}, Z) \\ B_t &= I(X_{2t}; Y_t | X_{1t}, Z) \\ C_t &= I(X_{1t}, X_{2t}; Y_t | Z) \\ D_t &= I(X_{1t}, X_{2t}; Y_t) \end{aligned}$$

$t = 1, \cdots, 9$. Equations (127) are satisfied. We then identify $\mathcal{R}_t$ of this appendix with $\mathcal{R}(P_{ZX_{1t}X_{2t}Y_t})$ of (60), $t = 1, 2, \cdots, K$, and $\mathcal{R}_0$ with

$\mathcal{R}^*(P^{(K)}_{Z\mathbf{X}|\mathbf{X},Y})$ of (59) which is consistent with (130). Then (129) and (131) yield (60). Q.E.D.

REFERENCES

1. Shannon, C. E., "A Mathematical Theory of Communications," B.S.T.J., *27*, No. 3 (July 1948), pp. 379–423, No. 4 (October 1948), pp. 623–656.
2. Shannon, C. E., "Two Way Communication Channels," *Proceedings of Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 1961, pp. 611–644.
3. Van der Meulen, E. C., "A Note and Counterexample on the Two Way Channel," Center for System Science, 70–15, University of Rochester, 1970.
4. Liao, H., "A Coding Theorem for Multiple Access Communications," *1972 International Symposium on Information Theory*, Asilomar, California, 1972. Also Ph.D. dissertation, "Multiple Access Channels," Dept. of Electrical Engineering, University of Hawaii, 1972.
5. Van der Meulen, E. C., "The Discrete Memoryless Channel with Two Senders and One Receiver," *2nd International Symposium on Information Transmission*, USSR, 1971.
6. Ahlswede, R., "Multi-way Communication Channels," *2nd International Symposium on Information Transmission*, USSR, 1971. To appear in Problems of Control and Information Theory.
7. Cover, T., "Broadcast Channels," IEEE Trans. Information Theory, *IT-18*, 1972, pp. 2–13.
8. Bergmans, P. P., "Random Coding Theorem for Broadcast Channels with Degraded Components," unpublished work.
9. Gallager, R. G., *Information Theory and Reliable Communication*, New York: John Wiley and Sons, Inc., 1968.